

R E P O R T R E S U M E S

ED 003 601

56

ACTIVITY LEVEL AND LEARNING EFFECTIVENESS.

BY- STAKE, ROBERT E. SJOGREN, DOUGLAS D.

NEBRASKA UNIV., LINCOLN

REPORT NUMBER NDEA-VIIA-753

PUB DATE MAR 64

GRANT OEG-7-37-0220-147

EDRS PRICE MF-\$0.18 HC-\$3.80 95P.

DESCRIPTORS- *LEARNING ACTIVITIES, *ACHIEVEMENT RATING,
*COLLEGE STUDENTS, *PROGRAMED INSTRUCTION, *INSTRUCTIONAL
AIDS, LINCOLN, NEBRASKA

A STUDY OF LEARNING ACTIVITY EXPLORED (1) AN ACTIVITY-ACHIEVEMENT SCALE TO DESCRIBE THE IMPACT OF ACTIVITY ON ACHIEVEMENT AND (2) THE POSSIBLE COMPLEXITY OR DIMENSIONALITY OF THIS IMPACT. TEN GROUPS, OF 10 COLLEGE UNDERGRADUATE STUDENTS EACH, WERE SCHEDULED TO STUDY UNDER EACH OF 10 LEARNING SITUATIONS. THE SITUATIONS CONSISTED OF TWO MODES OF STUDY (WRITING AND NOT WRITING) FOR EACH OF FIVE CONDITIONS--(1) NOTES, (2) OUTLINE, (3) TALK, (4) PROGRAMED TEXT, AND (5) TEACHING MACHINES. ONE TOPIC WAS ASSIGNED TO ALL STUDENTS EACH WEEK FOR 10 WEEKS. ANALYSIS OF VARIANCE IN ACHIEVEMENT SCORES SHOWED THAT THE DIFFERENT INSTRUCTIONAL SITUATIONS DID YIELD DIFFERENT LEVELS OF ACHIEVEMENT. LEARNING WAS MORE EFFECTIVE IN SITUATIONS WHERE PROGRAMED MATERIALS WERE USED. THREE POTENTIAL ACTIVITY-ACHIEVEMENT SCALES WERE TESTED TO DESCRIBE AND RELATE THE SEVERAL CONDITIONS. THE FACT THAT ACTIVITY WAS A CRITICAL VARIABLE IN DETERMINING ACHIEVEMENT WAS NOT PROVEN. A SCALE OF MEANS WAS THE BEST BASIS FOR GENERALIZING AS TO THE EFFECTIVENESS OF STUDY CONDITIONS. WITH ITS SCALE REFLECTING THE POTENTIAL OF EACH STUDY SITUATION FOR ALTERING INDIVIDUAL PERFORMANCE, AN IMPACT MODEL WAS FOUND TO BE NO SUBSTANTIAL IMPROVEMENT OVER THE CONVENTIONAL ANALYSIS OF VARIANCE MODEL. (JM)

ED003601

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

753
153

Activity Level and Learning Effectiveness

Robert E. Stake
Douglas D. Sjogren

UNIVERSITY OF NEBRASKA

MARCH 1964

Title VII Project Number 753
National Defense Education Act of 1958
Grant Number 7-37-0220-147

The research reported herein was
supported by a grant from the
U. S. Department of Health Education and Welfare
Office of Education

ACKNOWLEDGMENTS

The authors wish to acknowledge the advice and assistance provided by many individuals in the planning and execution of this project. Consultation on theoretical and psychometric considerations came from Ledyard Tucker, John Hunter, and Gene Glass. J. Thomas Hastings provided editorial advice.

Paul Thayer of Life Insurance Agency Management Association gave permission for the use of their life insurance program. Brent Baxter, Ray Hedberg, and Hans Steffan of the Prudential Insurance Company assisted with the insurance program. We also want to acknowledge the help of a kind lady in the Chicago Public Library, whose name is not known. She broke up a tea party to get us some books that we had to have for the Dewey Decimal program.

Programs were written by Lester Harvey, Ray Wiman, Gene Glass, Floyd Urback, and Helen Longstreth.

Graduate students who helped with the project at various stages were Neal Hartman, Tom Osborne, Neal Zimmerman, and Jan Cureton.

Audria Shumard typed the manuscript and secretarial services were provided by Marian Shemberg and Maxine Judd.

The authors are most appreciative of the help and advice provided by all of the above named individuals.

This research study was made possible by support from the Title VII program of the National Defense Education Act administered by the U. S. Office of Education.

TABLE OF CONTENTS

Chapter	Page
I THE ACTIVITY DIMENSION	1
Contemporary Issues in Learning	2
The Multidimensionality of Learning Activity	3
The Activity-Achievement Scale	4
Criteria for the Scale	5
The Choice of Dimensions	6
Set, medium, reinforcement	6
Rational Dimension One: Writing	7
Rational Dimension Two: Programing	8
Rational Dimension Three: Novelty	9
Objectives of the Research	11
II THE RESEARCH DESIGN	12
The Schedule	13
Lesson Objectives	16
Instructional Materials	16
Programs by Other Authors	19
A Kit of Programs for Experimentation	19
Materials Try-out	19
The Student Body	22
Instructions to the Learners	22
Measurement of Achievement	25
Other Measurement	26
Miscellaneous Information	26
Apparent Study Activity	27
Statement of Hypotheses	30
III UNIVARIATE ANALYSIS OF ACHIEVEMENT	33
Item-Type Bias	33
Pretest to Posttest Gains	35
Quiz Reliabilities	35
The expected Order of Means	39
The Observed Order of Means	39
Analysis of Variance of Achievement Scores	42
Components of Variance	46
The Mean Squares	46
Error Terms	49
Independence of Observations	49
Comparison of Pairs of Study Situations	50
Results of the Analysis of Variance of Achievement	
Scores	51
Retention	52

Table of Contents (Continued)

Chapter	Page
Personal Values and Reaction to Programs	56
Opinions on Instructional Materials	60
Attitude Toward Machines	63
Initial Equivalence of Groups	66
 IV AN ANALYSIS OF IMPACT	 67
An Emphasis on Individual Differences	67
Functional Scales	68
A Linear Regression Model	69
Errors of Fit	70
Solving for Parameters	71
Solving for Dimensions	72
Graphic Examples of Impact	72
The Distinction between Impact and Correlation Analyses .	76
Impact Analysis of the Achievement Data	76
 BIBLIOGRAPHY	 83

LIST OF FIGURES

Figure		Page
1	Schematic diagram of the 10 study situations	10
2	Three dimensional schematic of the experimental design. (Model A.)	14
3	Comparison of the apparent study involvement of students (W) making written responses during study with students (R) studying without making written responses	28
4	Comparison of the apparent study involvement of 160 students (P) studying programed materials with 240 students (C) studying conventional materials during 90 minute study periods	28
5	Percent of students appearing to be studying under 10 different study conditions during 90 minute study periods .	29
6	Unidimensional Activity-Achievement scale with means of 10 situations	41
7	Model B, a structural representation of the experimental design	43
8	Model C, a structural representation of the experimental design	44
9	Unidimensional representation of achievement by 4 learners in 3 conditions	73
10	Criterion performances of 4 individuals without interaction, 4 individuals with interaction but without multidimensionality, and 4 individuals with both interaction and multidimensionality	75
11	Unidimensional representation of achievement of 4 individuals	77
12	Four experimental conditions requiring 2 dimensions	77
13	Unidimensional representation of achievement in 4 conditions having equal means	78
14	Three achievement scales	81

LIST OF TABLES

Table		Page
1	Study Situation Schedule for Ten Groups of Ten Learners Each .	15
2	Titles and Composition of Lesson Materials	20
3	Factor Loadings and Communalities of Variables Sorted as to Item Type	34
4	Characteristics for 63 Achievement Quizzes	36
5	Topic Means on Pretest and Posttest and the "t" Values for the Difference	38
6	Means and Standard Deviation of Achievement Scores and Average Busyness Scores	40
7	Main Effects and Interactions	45
8	Components of Variance for Analysis of Variance of Achievement Scores	47
9	Complete Analysis of Variance of Achievement Scores	48
10	Summary of Analysis of Variance of Achievement Scores	50
11	Test and Retest Means for 100 Students Studying the First Five Topics	53
12	Analysis of Variance of Retention of Five Topics	54
13	Analysis of Variance of Loss Scores by Topics	55
14	The Correlation between Achievement and Personal Values as Found in Ten Instructional Situations	57
15	Mean Deviations in Achievement of Students with High Value Scores from Achievement Means for All Students	58
16	Numbers of Students with Various Value Scores Objecting to the Use of Teaching Machines	59
17	Numbers of Students Preferring Each of Ten Topics	60
18	The Topic Liked Most and the Topic Learned Best as Indicated by 100 Students	61

List of Tables (Continued)

Table		Page
19	Student Preferences for Topics Classified as to Conductiveness to Achievement For the Whole Group	62
20	Number of Students Preferring Topics Classified as to Response Type and Study Material Type	63
21	Changes in Student Reactions to Programed Materials During the Study	64
22	Group Means and "F" Values of 5-Item Pretest and Posttest of Each Topic	65
23	Reduction of Error Variance in Achievement Scores by Analysis of Variance and by Impact Analysis	80

CHAPTER I

THE ACTIVITY DIMENSION

We have heard that when John Dewey was asked about the statement, "we learn by doing," he snorted and allowed that this was one of those gross oversimplifications of the slogan-makers. Such was not his way of putting it. He had said that

"...power and control are obtained through realization of personal ends, personal selection of means and materials which are relevant, and through experimentation and testing" (1899).

He felt that the child is not "a listening being," and added,

"The statement....that education means 'drawing out' is excellent if we mean simply to contrast it with the process of pouring in. But, after all, it is difficult to connect the idea of drawing out with the ordinary doings of the child of three, four, seven, or eight years of age. He is already running over, spilling over, with activities of all kinds." (1900, p 37).

This research project concerns the activity of learning, and is indeed another attempt at oversimplification. In undertaking it we wanted to see if we could find a functional equivalence among a number of seemingly dissimilar study activities: "What different activities yield equal achievements?" Not only, "Is achievement enhanced more by one activity than another?" but also, "With the same achievement, are some persons more successful with one study activity, whereas others are more successful with another?" To look for functional equivalence among activities, (or systematic departure from equivalence), is to look for a simplification, a simplification that for some purposes would assuredly be an oversimplification. Yet the major advances in science have been achieved by unconventional and parsimonious descriptions of complex phenomena. Our research has been an effort to find a new and simple way of analyzing and describing the complex activity of learning.

If, from an analysis of learner responses, a large number of activities are found to facilitate learning, but some to different degrees than others, we can speak of an Activity Dimension. If the effect of different activities is not consistent across learners, we will speak of the interaction of learners with modes of learning. If an individual is consistently influenced by certain instructional conditions, we will speak of the susceptibility of this learner. If a

certain activity is found consistently to influence individual achievement, we will speak of the impact of this activity. The impact of different activities can be scaled. If we find that we can explain achievement better by referring to several scales or continua or bases for instructional conditions, we will speak of the multidimensionality of learning activities. This report is a summary of research on the dimensionality of certain activities of learning.

Contemporary Issues in Learning

As psychologists and educators have studied the learning process during the twentieth century a number of major issues have emerged. There is considerable disagreement even as to what the learning process is. To some, learning is a process of perceptual reorganization. The critical task for the educator is to confront the learner with stimuli in such a way that he may augment and restructure what he knows. (Koffka, 1935; Lewin, 1942; and Tolman, 1934) The main alternative position is that learning is a process of modifying responses. Since different responses have a different likelihood of occurring, the educator's job is to increase the likelihood that desired responses will occur and to decrease the likelihood of occurrence of other responses. Learning occurs when a learner associates; behaviorally, not necessarily consciously; the desired response with a certain configuration of stimuli. (Hull, 1943; Thorndike, 1935; Gates, 1942). But the emphasis is on responses rather than on the stimuli.

It is probably obvious that the two viewpoints of this perception-response issue lead to quite different emphasis on learner activity. As Deese and others pointed out, (Meierhenry, 1961) the Gestalt-oriented emphasis on perceptual reorganization leads to the use of multisensory aids for instruction, to capture the learner's attention and to enrich his concepts; and the stimulus-response-bond viewpoint, on the other hand, leads to a structuring of the learning environment to permit more participative learner behavior and to elicit specific behavior.

A second major issue is whether reinforcement is a necessary element or whether it is merely a facilitating element in the learning process. In either case, there is agreement that learning can be accelerated and extended and refined by use of incentive. The educator should be aware that one cannot control the schedule of reinforcement unless the learning activity is overt, or manifested in some subsequent overt activity, such as recitation (Skinner, 1954).

An issue particularly important to some of the practitioners of education has been the matter of whether to organize lessons around the whole task, or to divide the whole into parts to be learned one by one. Much of the research on this whole-part problem has been done in the laboratory with lists of words and nonsense syllables (Hovland, 1951). It is a live issue today because auto-instructional materials are being developed for the most part by programmers who break the subject matter

into small parts. This is done, not because it has been found that the part method is better than the whole method, but because these programmers desire to provide frequent opportunities for learner responses. A holistic approach and frequent responding are probably not incompatible but at the time of this writing the emphasis is on elementalism and reinforcement. The activity of the learner is influenced, of course, by whether his learning tasks are presented in wholes or in parts. In the former case, the more covert acts of contemplation are being emphasized, whereas in the latter, the more overt responses are being elicited.

A fourth and last issue to be cited is the issue of closure. Those psychologists who have been identified with the Gestalt point of view have emphasized that we humans strive to place our thoughts and our perceptions into balance, so that what remains is logical, harmonious, complete. We attempt to gain a sense of closure. Zeigarnik (1927) raised an important issue when she found that a task for which there is closure is not as well remembered as the task that remains unsolved or unfinished. Recently, Briggs, Goldbeck, Campbell, and Nichols (1962) have observed a superiority in retention by learners responding covertly, without feedback, to programmed materials; and they have suggested that this result may in part be attributable to the Zeigarnik effect. The learner who makes no overt response may in fact lack the closure that comes with a written commitment. Most educators would be aghast at the possible implication that you should lead a student up to the solution of a problem but delay the attainment of the solution.

More will be said of some of these issues later in this report. The purpose of introducing the issues here was to support the contention that many of the important research questions in learning are related to the activity of the learner. We should be better able to deal with these issues if we can improve our understanding and control of the acts which constitute or which accompany the learning process.

The Multidimensionality of Learning Activity

From a rational point of view, the activity of a learner is variegated. Various motor actions are apparent and various mental actions can be presumed. The behaviorist in us observes body orientation, eye movement, facial expression, trial and error, vocalization, quiescence, and acquiescence. The phenomenologist in us observes induction, attention, extrapolation, judging, wishing, and loving. All of these one can call "acts," as Brentano did (1911). In terms of our traditional perspectives, learning is not a single act but a myriad of acts, i.e. a multidimensional activity. A learner can be more, or less, active in terms of any of these dimensions. Some of his activity influences the effectiveness of his learning. We are aware that increased activity sometimes facilitates the learning, but it is obvious

that different activity is appropriate in different learning situations and that too much of some activity can interfere with learning.

When these workaday conclusions are drawn it is rather customary for us to say something like, "Well, what specific acts should be increased if we wish to facilitate learning of this specific task by those particular learners?" But this is not the major question to be considered here. Rather different questions are to be considered: questions concerning the dimensionality of learner activity as manifested in achievement scores. Are achievements structured multidimensionally? Is the manifest structure consistent with the rational structure our theorists have built? Such are the questions being investigated.

The Activity-Achievement Scale

Although we perceive learning to be a multidimensional act, it would be advantageous if, for some purposes, we could discuss learning with reference to a single scale of activity. Fry has cited over 200 variables in programmed instruction (1963). To discuss such instruction we must ignore some of the variables or reduce the many to a few basic pervasive variables. Is it possible to reduce the different activities to a single activity scale? The problem of equating a unit of eye movement with some unit of vocalization with some unit of emotional feeling is every bit as perplexing and arbitrary as the problem of measuring some of these phenomena. One way to express the measurement of different activities in terms of a common scale would be to measure their different influences on learning and (if there is some consistency of influence) to construct a single activity-achievement scale.* Any particular amount of activity along a given dimension, when other activities are continued at a constant level, could be represented by a point on the activity-achievement scale. Reference to such a scale would be valuable in both theoretical and practical affairs.

As an example we might take the instance of the 6-year-olds memorizing the pledge of allegiance to the flag. We might be interested in two dimensions of activity, oral expression and emotional involvement. Let us presume that we have two levels of oral expression; X and Y, and three levels of emotional involvement A, B, and C. If an activity-achievement scale were justified we could represent the six

* It is possible that there would be a unique activity-achievement scale for each task and for each set of learners but it can be assumed for the time being that there would be considerable correspondence among many of the scales and that a single scale would have some general utility.

different activity conditions as points on a continuum as such:

AX	AY	CX	CY	BX	BY
----	----	----	----	----	----

Although C may represent greatest emotional involvement, the achievement with condition B is greatest. It is apparent from this scale that $B > C > A$ and that $Y > X$, both consistently. Furthermore, the difference between the two oral expression conditions that was so prominent with emotion condition A is somewhat diminished with emotion condition B.

If such a parsimonious representation were justified, the development of it would be of interest both to the experimentalist and to the methodologist. The teacher might find the idea of the scale helpful in planning his assignments. The administrator could discuss the costs of different study-environments in terms of the relative achievements. It seems at the time of this writing, however, that the immediate utility of the scale would be more in the realm of concepts and communication than in the realm of instrumentation and measurement.

The above representation shows a difference in activities with only a slight interaction between activities. Later in this report, the focus will be on interactions between persons and activities, particularly interactions which can be considered orderly, systematic, and functional.

Criteria for the Scale

In its simplest form an activity-achievement scale is unidimensional. It is a small matter to construct such a scale. The mean achievement for each activity level could be computed and those means would define the scale. The wisdom of such a choice of definitions depends on at least two matters. Can we be confident that there is a simple relationship between these activities and learning? And can we be confident that there is some difference between those means? The latter question is an analysis of variance question. The other question is a scaling question. A new scaling technique called impact analysis is presented in Chapter IV. Impact analysis was used to study the dimensionality of our learning activities.

If the performances associated with different activity conditions are found in essentially the same order for all learners, then the notion of a unidimensional activity-achievement scale is supported. If, on the other hand, at least two different orders are found repeatedly, the unidimensional activity-achievement scale would lead to an overly simplified description of the learning phenomena, and should be modified or abandoned for that learning situation. This examination of orders is accomplished as part of the scaling procedure.

What has been said is that the possibility of using an activity-achievement scale can be explored with analysis of variance and impact analysis. If significant differences and unidimensionality of the learning scores are found, a single scale could be used as a graphic representation of the activity of learning.

The Choice of Dimensions

In this research project we selected three possible dimensions of learning activity that are of special interest to the educator. We devised activities that might be a part of a typical scholastic assignment. To make this study relevant to much current research in educational psychology we chose to work with programmed learning and teaching machines, as well as with more traditional instruction.

Set, medium, reinforcement

We wish to point to three general ways in which teachers influence the activity of a learner. Preceding the learning, through personal interaction, verbal communication, and a command of prerequisite experience, the teacher influences the response set of the learner. Robinson (1946), Harlow (1949), Postman and Senders (1946), among others, have studied and written about the response set in learning.

The media that educators choose have much to do with the activity of the learner. Motion picture and television instruction draw attention but are not conducive to spontaneous discussion. Programed textbooks and teaching machines elicit certain responses but they inhibit such responses as scanning the material to get an overview of the lesson and dog-earing a page for quick future reference. By structuring the learning environment, particularly the lesson materials, the educator influences the activity of the learner.

Controlling and directing the activity of the learner through choice of instructional media has been discussed by Skinner (1954) and Deese (1958). There is some problem in distinguishing between control of response set and control of media. Modification of instructional approach or of materials will probably change the response set. An operational distinction can be made in terms of when the influence is exerted; if "before" the learning sessions begin, the modification is a modification of response set; if "during" the learning sessions, the modification is a modification of (or within) media.

The third way the teacher has of influencing response activity is through the use of reinforcement. By rewarding or by punishing or by providing for some form of self-evaluation the teacher attempts to strengthen some responses and to inhibit others. The reinforcement may be said to result also in change of response set, but it differs from set and medium in that it occurs after the response is made. The

relationships between reinforcement and learning have been discussed by Thorndike, (1927), and Postman, (1947), and many others (Jones, 1953-1963; Stone, 1948).

In selecting a set of learning experiences for this study it was desired that all three of these general determinants of activity be considered. The design might have been best if all three were varied systematically, but with the time and funds available only a few situations could be studied. We could have devised a 2x2x2 design, using all eight combinations of two response sets, two media and two reinforcement levels. Relatively personal and impersonal response sets could have been evoked. Reading and writing responses (media) could have been required. Knowledge of success in responding could have been provided for some and withheld from others as variation in reinforcement. Thus the three different influences on scholastic activity would have been included.

We chose to proceed on a slightly different tack. We chose to focus on several media of teaching that are widely used today, but while still attending to differences in response set and reinforcement. The three dimensions of activity that were selected as rational dimensions for this research are directly related to set, medium, and reinforcement. Experimental limitations permitted but about 10 different study situations (activity conditions). Many response sets, many media, and many schedules of reinforcement, of course, had to be ignored.

In this report we want to distinguish between the dimensions that are based on what we believe to be important, on the one hand, and the dimensions that can be extracted from an analysis of the learners scores, on the other hand. We will call the dimensions believed to be important, rational dimensions; and the dimensions obtained from quantitative analysis, performance dimensions.

Rational Dimension One: Writing

There continues to be interest and doubt in educational circles about the value of overt recitation during study. A typical way to obtain overt recitation is to instruct the student to make notes, to outline or to answer questions during study. Gates studied the effects of devoting various portions of an individual's study period to recitation. He found, especially with disjunctive material, that the more time spent in recitation, the better the learning--even up to 80% of the time (1917).

In the last four years, there has been considerable investigation of the value of an overt response to programmed materials. Silberman has reviewed a number of relevant studies. He reported that significant differences in achievement were found in almost half the studies but that sometimes the difference favored the covert response and sometimes it

avored the overt response (1962). Deterline (1962) and others have discussed this issue. The consensus seems to be that covert responses are likely to result in better achievement scores when the task is easy for the learners, when the task requires a very long study session, or when there is a time limit such that few students can cover all the material. It is recognized that such results could be entirely due to two things: namely that overt responding takes time and energy. When the task is easy the learner can master it with almost any approach, and he may become bored and discouraged by the expenditure of energy required by the overt method.* When the task is very long, discouragement or fatigue is likely to take a greater toll on the overt-responding learner. When the time limit prevents a coverage of all the subject matter, the covert learner probably has a chance to encounter more of it.

The conditions under which a written response yields substantial improvements are not yet known. It should be acknowledged that the difference between a written response and no written response is not necessarily the difference between a response and no response. The learner is capable of covertly acting out most of the elements of an overt response, we think, and if he chooses to do so, we should not anticipate much of a difference between a requirement for written responses and a restriction on written responses during learning.

Because of the interest in this requirement, because of its relevance to schoolwork, and because of its ease of administration, the writing of responses was chosen as one rational dimension for this investigation.

Rational Dimension Two: Programing

As is implied in the previous paragraphs and as the reader is surely aware, there is extensive contemporary interest in programmed learning. A high proportion of current educational research involves this instructional development. It was realized, however, that to include programmed materials in this study would introduce, simultaneously, a number of important variables. In including them, certain precision and control would be lost. In spite of this, the contemporary interest seemed to justify the consideration of programmed materials.

What are the variables that are introduced with programmed materials? More specifically, what are the variables of linear programs? These programmed materials differ from conventional materials in at least four important ways. The programmed materials are arranged in

* A colleague, Neil Hartman, has reminded us that Hull (1943) spoke of Reaction Inhibition. Attributable to the effort expended during learning, this inhibition was said to interfere with the behavior learned.

a different format, particularly in that they have been fragmented into small steps, almost automatically resulting in the teaching of parts rather than wholes. With the programed materials there is internal provision for specific learner responses and for knowledge of results, a valuable form of reinforcement. Lastly, at the present time the programed materials are novel for most learners and, at least at the outset, command the attention of our curious students.

What acts occur when our students use programed materials? Two kinds of acts are obvious. As in any learning environment, there are acts associated with the subject matter and acts associated with the physical materials. One of the most obvious activities, particularly of the novice, is the manipulation of paraphernalia. At first, at least, the expensive machines are distractive and the inexpensive machines need coaxing. With some devices and with some of the learners, fumbling vanishes along with the cues. But with others it does not seem to.

An active confrontation, the learner with his subject matter, is very apparent with the programed materials. Generally it is hard to doubt that a learner is using his best mental powers, if needed, in studying the frames of these lessons. In our work we have found that programed materials seem to bring about a higher level of gross activity. In spite of the problems of confounding the analysis of results, the activity of studying programed material was included as the second rational dimension.

Rational Dimension Three: Novelty

The two dimensions involving writing and programing might have satisfied the investigators in this initial attempt to justify an activity-achievement scale, but the difference between conventional materials and programed materials as stated a few paragraphs back loomed large. It was decided that another dimension would be added, a dimension involving another form of novel assignment. Conventional materials would be used in some relatively unusual assignment. Of course, a danger with novel assignments, including programed learning assignments, is that the students will attend to the novel conditions and not to the intended task. Higher achievement may be the result of the Hawthorne Effect (Cook, 1962) The activity of a novel assignment may be at cross-purposes with the activity needed for learning.

It occurs to most teachers that they really (finally) learn something thoroughly when they are obligated to teach it. The novel assignment chosen for this investigation was an assignment to prepare oneself to explain orally the lesson to his peers. It was not to be known by him whether or not the explanation would actually occur, but he was directed to prepare himself. Only conventional materials were used with this assignment.

Thus the third dimension was chosen. The conditions are schematized in Figure 1, with each block representing one of the study situations.

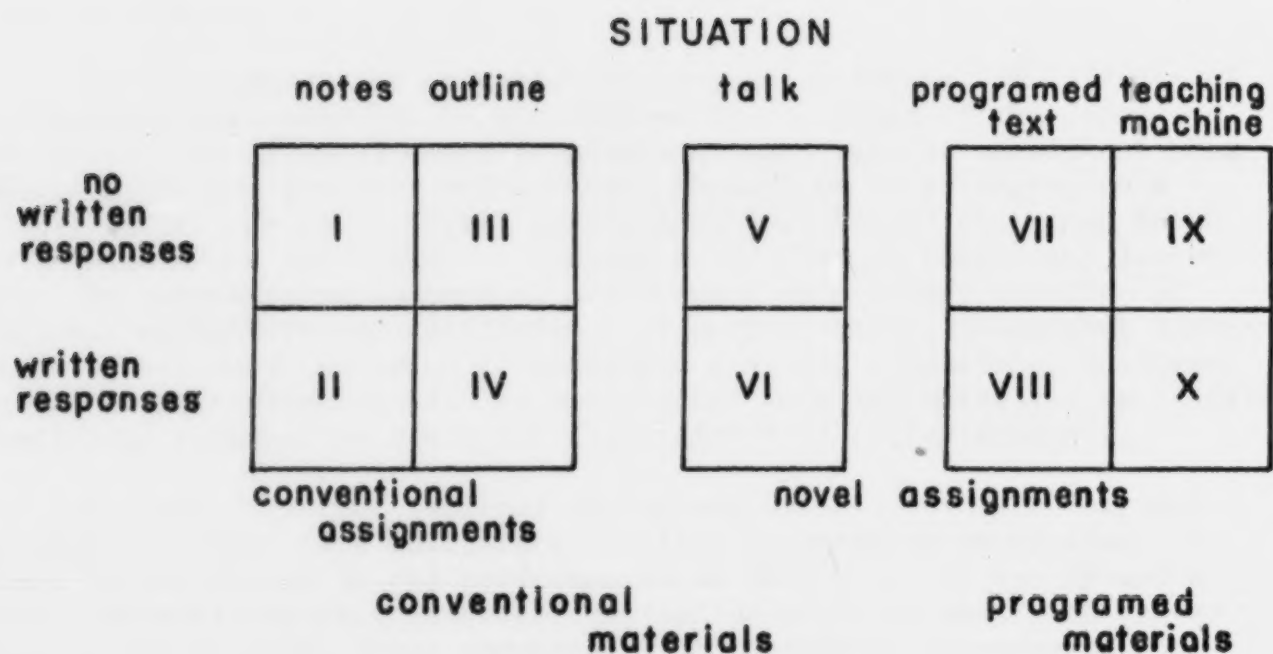


Figure 1. Schematic diagram of the 10 study situations.

Situation	Instructions
I	Read conventional materials, make mental notes.
II	Read conventional materials, make written notes.
III	Read conventional materials, make a mental outline.
IV	Read conventional materials, make a written outline.
V	Read conventional materials, plan a talk without notes.
VI	Read conventional materials, plan a talk with written notes.
VII	Read programed materials, make no written responses.
VIII	Read programed materials, make written responses to each frame.
IX	Read programed materials, make no written responses.
X	Read programed materials, make written responses to each frame.

Objectives of the Research

The major objective of this research was the study of learning activity. Our interest was both in the impact of activity and in the possible complexity or dimensionality of this impact. At the outset we hoped to devise a technique for describing the impact of these activities on achievement.

In the simplest of comparisons, here as elsewhere, the effects of independent variables can be represented by the means of the dependent variables. An array of means is unidimensional, so the conditions themselves might be, possibly erroneously, thought of as belonging to a simple order. In our analysis we hoped to look beyond the group means, to inquire about the effect of various activities on individual learning. By representing individual achievement as a linear function of personal parameters and coefficients of certain basic, underlying dimensions of activity, we hoped to provide a scale of conditions, unidimensional or multidimensional. We anticipated enabling ourselves to classify conditions somewhat as tests are classified in a factor analysis.

Several different rational dimensions of activity were introduced so that, if there is a multidimensionality of learning conditions, it might be manifested in the performances of learners. It was deemed a likely possibility that adequate description might be made with reference to but a single basic dimension with conditions represented not necessarily in the order of their means. (The placement of activities along this continuum would be an indication of the degree to which each affects individual performances). Second, third, or even further dimensions might be needed, however, to provide the basis for our description of learning.

We measured the achievement of each member of a group of learners in each situation shown in Figure 1. Achievement scores were processed with an analysis of variance and were scaled to determine if the ten instructional situations could be meaningfully and adequately represented by ten separate points in a unidimensional space. It was recognized that failure to accomplish this end could be due to a lack of confidence in the differences of impact on achievement for the ten conditions, or due to the multidimensionality of performance resulting from systematic differences in learners reaction to the different study situations.

The principal objective of this research was the exploration of the activity-achievement scale. Collateral objectives are discussed in the next chapter.

CHAPTER II

THE RESEARCH DESIGN

It is our contention that one of the major obstacles to further understanding of human learning can be traced to inattention to the differential effects of aptitude, experience, aspiration, and perseverance: those personal characteristics. In the usual experimental design, individuals are sampled randomly for each treatment or combination of treatments. Since no one is subjected to more than one set of experimental conditions, important systematic differences between individuals cannot be estimated separately from the effects of quotidian variation in performance, errors of measurement in instruments, and all other vagaries of observation. Since the effects treatment must be a certain stochastically-determined multiple of the effects of error in order to be acknowledged, the lumping into error of variance due to systematic personal differences requires the effect of treatments to be unnecessarily large. Under these circumstances meaningful differences and relationships are likely not to be distinguishable from the effects of random error. Such Type II errors can be lessened by exposing the same learners to all treatments. If this can be done, the variation due to systematic individual differences can be estimated apart from all other forms of experimental error. The result of these repeated measurements of learners is increased precision for the analysis of the data with greater likelihood of discerning treatment effects of a given size. In this study we decided, as one of the first orders of business, to involve the same learners in all activity conditions.

There are two serious problems that were introduced by this early decision. One problem involves the interpretation of an analysis of variance when measurement cells are not independent. This problem will be discussed briefly in the next chapter.

The second problem is reflected in the question, "What task or topic can the students be given in the second study situation after they have studied something in the first situation?" If the two topics are interdependent, the investigator needs to have very effective tests to determine how much was learned under each of the two situations. Lacking such tests, we concluded that the subject matter for each of the study conditions should be carefully chosen so that even complete mastery of one topic would not give a student an advantage on any other topic. The choice of topics will be discussed at greater length in one of the following sections.

Since each student was to study under each of the study situations the group had to be divided into as many subgroups as there were situations. Ten situations had been selected, as indicated in Chapter 1, so ten groups of learners were obtained and ten lesson topics were chosen.

In harmony with this decimal motif, ten learners were assigned to each group.

It was realized that some of the students would be acquainted and would discuss their activity with each other. We didn't want them to learn anything relevant outside the experiment. This talk might have been minimized if each group had been working on a different topic at any one time. On the other hand, the sessions had to be interrupted by a vacation period. To have different groups studying different topics would have meant that for each topic, some students would be tested after two weeks, others after three. This was considered less desirable than the possibility of the discussion of topics when all groups studied the same topic each week.

Let us summarize the experimental design. Ten groups of ten students each were scheduled to study under each of ten learning situations. The situations were divided into two groups of five conditions: notes, outline, talk, programed text, teaching machines. With each of the conditions the students used two modes of study: writing and no writing.

A new topic was studied each week in a new learning situation. The placement of group A is shown as a shaded block in Figure 2. Block J is sliced into ten levels to remind us that there are 10 learners in each group.

In terms of Lindquist's typology of designs (1953, p 302), this model has no numerical designation. With A, B, and C representing the classifications, Mode, Condition, and Topic, $a = 2$, $b = 5$, and $c = 10$. Ours is a special case of a mixed design where $ab = c$. The main effects and AB interaction are "within" effects, i.e. administered to all students, but the AC, BC, and ABC interactions are "mixed" effects, i.e. several but not all combinations administered to each student. These mixed effects are not evaluated as effectively as the within effects, but we were not very interested in interactions with Topic in this study anyway.

The Schedule

One topic was assigned to all students each week. The "students-studying" period, then, extended over ten weeks. A prior week was used for orientation, another week for spring vacation, and two weeks for final testing, making a total span of fourteen weeks.

The assignment of groups to study situations is shown in Table 1. The Roman numerals refer to the study conditions listed in Figure 1. Groups were roughly paired according to avowed previous acquaintance with the ten topics. The 5 pairs were assigned to the 5 conditions randomly for the first week, then rearranged for the next four weeks to conform to a Latin Square design. During all of the first five weeks

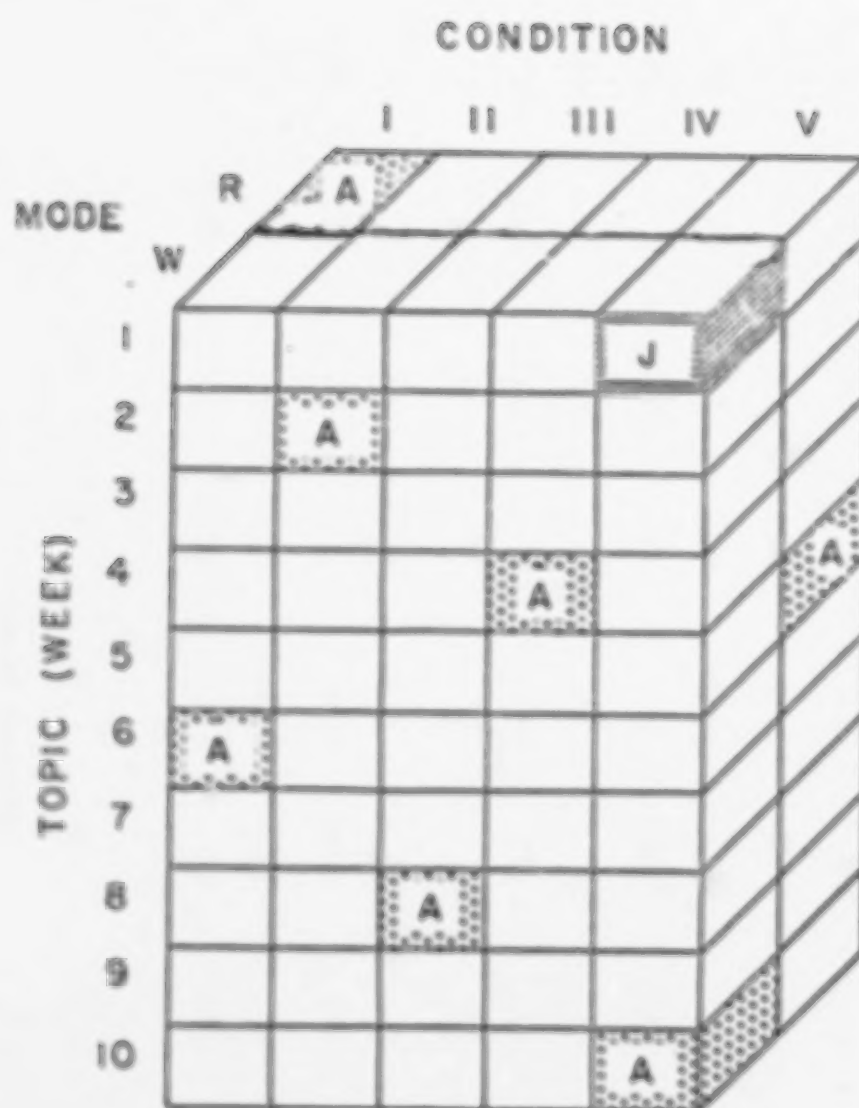


Figure 2. Three dimensional schematic of the experimental design. (Model A.)

Table 1

Study Situation Schedule for Ten Groups of Ten Learners Each

Week	Learner Group									
	A	B	C	D	E	F	G	H	I	J
1	Orientation									
2	I	II	III	IV	V	VI	VII	VIII	IX	X
3	IX	X	I	II	III	IV	V	VI	VII	VIII
4	V	VI	VII	VIII	IX	X	I	II	III	IV
5	III	IV	V	VI	VII	VIII	IX	X	I	II
6	VII	VIII	IX	X	I	II	III	IV	V	VI
7	II	I	IV	III	VI	V	VIII	VII	X	IX
8	X	IX	II	I	IV	III	VI	V	VIII	VII
9	VI	V	VIII	VII	X	IX	II	I	IV	III
10	IV	III	VI	V	VIII	VII	X	IX	II	I
11	Vacation									
12	VIII	VII	X	IX	II	I	IV	III	VI	V
13	Testing									
14	Testing									

Note.--A different topic was used for all learners each week.

one of the paired groups followed the no-writing mode. Its pair-mate was assigned the writing mode for all five weeks. In the final five weeks the order of conditions was repeated but with the opposite mode.

This Latin Square schedule balanced the order of situations. Half of the students encountered Situation I before Situation X, the other half after; etc. Each weekly session was scheduled for two hours, with 90 minutes of study and 20 minutes of quiz. Except when spring vacation made it three, two weeks elapsed between the study of a given topic and the quiz on that particular topic.

Lesson Objectives

A conclusion that has been drawn by readers of programed learning research is that statistically significant differences are not usually found with brief lesson materials. Some workers have expressed alarm at the possibility that the programed learning movement will be evaluated in terms of the short-term studies. Several urged the authors to abandon plans to use 90 minute materials in this study.

It would be unfortunate if the value of programing instruction were to be estimated only on the basis of what achievement we can measure after a brief study period. On the other hand, the particular value of brief, occasional, programed lesson supplements cannot be evaluated by studying semester-long instruction. The major purpose of this research was not to assess programed instruction, short-term or long-term, but to evaluate certain activities of learning. Programed materials were used because they evoke one bunch of activities. We chose to involve each learner in each activity in order to control irrelevant individual differences. It then was not possible to use lesson materials which required more than an hour or two of study.

Although all too little can be learned in 90 minutes it was believed that a valid quiz could be developed for each topic. With the controls of a Latin Square design, the quiz scores were expected to be sensitive to whatever important differences there were in our study conditions. The content of the quizzes is discussed in a later section.

Instructional Materials

The research design required ten different study topics. The following specifications were established for the selection of topics.

1. Self-contained. A topic which could be undertaken with approximately equal ease by any college-level student; a topic having no prerequisites other than what entry to college would presuppose.

2. Non-academic. A topic which most college students are unlikely to encounter in formal study.

3. Informational. A topic with which sufficient acquaintance could be developed in a brief period so as to permit discriminative achievement testing. .

4. Palatable. A topic that would be of some interest to most college students and which would have sufficient "dignity" so that the student would probably not see himself as "wasting time," in the study of it.

For each topic there was a need for lesson materials in both programmed and conventional format. Originally, it was intended that there would be both linear and branching programs but resources of the project limited the programing to the linear kind. Additionally, it was intended that programing would be such that different mental processes, (e.g., memory, inductive reasoning, and artistic appreciation) would be evoked in the study of the different topics.

There were specifications for the program format. Because the study-time was to be carefully controlled and because it was desired that the materials be arranged to minimize differences between fast and slow readers, both conventional and programmed materials were to follow a spiral development. The main content of the lesson was to be covered in initial sections, with later sections devoted to amplification, application, and implication. Thus, a slow reader who could thoughtfully project or intuit these extensions of the basic material would be little handicapped by his reading speed. With spiral presentation of information, differences in reading speed between learners studying in one situation and learners studying in the others are probably going to be of diminished importance.

The length of the study period was set at 90 minutes. All students were required to participate the full 90 minutes. It was hoped that the conventional materials would require slightly less time than the programmed materials for most students, with the remaining minutes available for review and summary. Review frames were provided as an extension of the programmed materials and made the programs long enough so that few students completed all frames. As indicated in the previous paragraph, the frames not reached by a slow reader were supposed to contain only amplification of information previously encountered.

Much contemporary academic programing is too self-contained (Stake, 1963). Little or no attention is given to developing competence with reference publications. Here, at least several programs were written to involve resource publications or even to be vehicles for learning the use of a reference publication. In another departure from common programing practice, an occasional, difficult frame was included. This was done to weaken any possible response set for making only well-learned responses.*

* Introspecting students have reported that with the 95% difficulty materials they get to looking for key words in the frame to copy and

During the research planning stages it was presumed that programmed and conventional materials could be obtained to meet these many specifications, but such was not found to be the case. Most programs could not be used in their entirety because of study-time limitations. The introductory sections of most programs were written to pursue attitudinal or study-method objectives and so contained too little content for testing. Subsequent sections, if the program was a good one, required previous study of the first section. Some sufficiently short programs were deemed to be unpalatable for the learners here.

The Topics

Originally it was planned that topics would be selected in non-sequential pairs in order to get a parallel-forms indication of reliability of the achievement. Such topics as Economy of Bolivia and Economy of Ecuador, Insurance and Investments, the English Money System and the German Money System were considered but only the latter pair was ultimately included.

In the other four pairs, the pairing was based on similarity of learning process more than similarity of content. Two topics Insurance and Nebraska State Government were detail-oriented subjects, mastery of which was largely attainable by rote memory. The topics, Dewey Decimal Classification and Leaves, were both believed to require an exercise of inductive powers. The Poetry and Art lessons were more analytic than aesthetic, but some attention was given to artistic feeling. The German Money and English Money materials called for little more than rote memory and some competence in arithmetic. The lessons on Logic and Fermentation probably involved the most systematic content demanding an understanding of relatively complex relationships. The experimenters were more satisfied with their success in tapping different mental processes with this selection and development of lesson materials than with their success in devising programs with parallel content.

When it became apparent that we would not find lesson materials meeting the specifications indicated earlier, we decided to write and assemble our own. Several colleagues indicated a desire to write a suitable linear program but only two succeeded in completing even a first draft. These persisting scholars were Floyd Urbach, who wrote or

for key words most recently taught. All frames answerable at the 95% level are not oriented to simple recall of key words, of course; but a set toward keyword-search can probably be diminished by a lessened emphasis on technical vocabulary and by frequent, difficult inquiry. Perhaps the rhetorical question has just as much place in programmed material as elsewhere.

assembled all the materials on fermentation, and Ray Wiman, who programmed material on classification of leaves. Numerous graduate students participated in one job or another, e.g., writing, assembling, editing; but the brunt of the work was borne by the authors of this report.

Programs by Other Authors

Programed materials on life insurance were made available by Hans Steffen of the Prudential Life Insurance Company and by Paul Thayer by the Life Insurance Agency Management Association. David Klaus of the American Institute for Research permitted us to use his program on the English money system as part of our material for that topic. Similarly with the program on logic, we were able to incorporate a good bit of material prepared by J. W. Blyth and J. H. Jacobson, Jr. The poetry program was an adaptation of a program developed by General Programed Teaching Corporation. (Naturally they retain jurisdiction over all future use of their programs. This is indicated by Note b in Table 2).

A Kit of Programs for Experimentation

It was the intent of the authors to develop these ten sets of materials so that they could be used as a laboratory kit by other investigators. This seems less likely in retrospect because so much of the material is written for use with a particular reference book and/or set of illustrations. The materials on poetry and art were developed for use with paperbound texts. Classification of leaves was written with sustained reference to a Handbook of Nebraska Trees, and the Dewey Decimal System is taught by using one edition of its primary reference manual. The two money lessons were devised for frequent reference to a pair of illustrated plates. In order for others to use these standardized programs a supply of the reference material would have to be obtained. For the benefit of potential users a table of details of the lesson materials is included as Table 2. Programed frames have been mimeographed on 8 1/2 x 11 paper. Indication is given when permission from other than the authors of this report is needed for duplication and use.

Materials Try-out

The programs and conventional materials were tried out with eight students from a local school of nursing. The students were asked to work the programed materials or to study the conventional materials as though they were preparing for a test over the content. When they had finished studying the material a preliminary form of the achievement test was administered. A record was kept of the time spent in working through the program or in studying the conventional materials. The students indicated sections of the material that were ambiguous or

Table 2

Titles and Composition of Lesson Materials

Number	Topic	Conven- tional pages ^a	Mimeo Frames	Panels	Supplementary Publications (one per student)	Outside Permission Needed
1	Life Insurance	14	258	1 mimeo	None	Yes
2	Dewey Decimal System	1	219	2 mimeo	Dewey Decimal Classification and Relative Index, Standard 15th Edition Revised	No
3	English Money	12	209	British Currency and Equi- valents (Barclays Bank)	English Money (Klaus, 1961)	Yes
4	Poetry Analysis	14	219	1 mimeo	The Pocket book of Modern Verse, Revised Edition, 1958	No
5	Rules of Logic	12	292	None	None	Yes
6	Nebraska State Government	1	336	1 mimeo	Nebraska State Government (League of Women Voters, 1961)	No

7	Leaf Types	5	188	7 mimeo	Handbook of Nebraska Trees. (Pool, 1951)	No
8	German Money	10	176	Kursfahige Bundesmun-zen (Deutsche Bundesbank, 1959)	None	No
9	Art Analysis	13	185	None	Composition, The Anatomy of Picture Making. (Pitman, 1958)	No
10	Fermentation	9	237	1 mimeo	None	No

^aSingle spaced, narrow margin

unintelligible and detected many typographical errors. The materials and tests were revised and shortened or lengthened on the basis of the responses of made by a small group of student nurses. No special attempt was made to conform to a set standard of performance, such as "95% correct," on the frames or the programs. The materials were revised, rather, on the basis of whether an error was judged to be the fault of the program. In most cases one try-out was made, but in a couple of instances the revision was quite extensive and the nurses "tried-out" the revised materials. The insurance and English money materials were revised twice.

The Student Body

Students were obtained by advertising the opportunity and \$25.00 pay. Of the almost 200 college undergraduates who applied 100 were selected and assigned to the ten study groups. The applicant was asked to indicate his familiarity with the ten lesson topics. Although the information was not validated in any way before students were assigned to groups, it was used in an attempt to equate the groups as to familiarity with the different topics.

A high majority of the students were freshman and sophomore girls enrolled in the Teachers College of the University of Nebraska. After the 100 were selected they took a 40 item test, 4 items per topic, as a basis for checking on the equivalence of the groups. The subtests were of low reliability, of course. Nothing was found to suggest that some groups were quite familiar with some topics whereas other groups were quite familiar with other topics. It was apparent that students in a couple of the groups were more mature and experienced and earned higher scores, but this was a general rather than a specific superiority. It was expected that such differences would balance out. Scheduling restrictions prevented more than a few reassignments, but the equating was believed to be satisfactory.

For the \$25.00, students agreed to commit a full 120 minutes each week for 14 weeks. It frequently happened that a student was absent, but his session was made up, usually within two days. On rare occasion it was made up as much as a week late. However, each of the 100 students participated in all 14 sessions in the order assigned his group. There were no dropouts.

Instructions to the Learners

Each weekly session with the students lasted for two hours. The study portion required the first 90 minutes. It was interrupted by a one-minute break about half way through. Twenty of the remaining minutes were used for testing. It had been decided that scheduling the study session first would tend to minimize the individual differences in review just prior to the tests. It was also felt that if the study

occurred during the last 90 minutes there would be a stronger tendency for some to extend the study period at home. The students were asked to refrain from such outside study.

The instructions given for each of the study conditions are listed below. Again, the Roman numerals refer to the situations listed in Figure 1.

I. Reading Only. "Your special study condition this week is to study the materials simply by reading them. You are not to make any notes of any kind. Just read the materials carefully and think about them. Put your pencils and note paper aside for the time being. If you read all the way through, go back over them again. Are there any questions?"

II. Written Notes. "Your special study condition for this week is to study the materials and to make a good set of notes. Suppose that you were going to want to review this material several months from now. While you read this material, jot down some of the information that you feel is most important. Jot it down in such a way that when you look at the notes several months from now, it would be easy to remember the information.

At the end of the study session today I will collect the notes. We will look at them to see that you did take notes, but we are not interested in grading the notes. The notes are supposed to be study-aids, not papers to be handed in to an instructor. Try to cover all the material in the 90 minutes. If you finish, go back over it again. Any questions?"

III. Mental Outline. "This week your study condition is to read the materials and to outline them mentally. You are to organize and structure the reading materials in your mind, without writing anything down. Of course, it is very difficult to keep a detailed outline in mind, so you will want your outline to be very simple. It might be best to read all of the materials or, at least, most of the materials, before you even start the outline. The most important thing in this study condition is to get an outline of the material well in mind. Do you have any questions before we begin?"

IV. Written Outline. "This week you are to prepare a good written outline on the study-materials. I will hand out the materials in a few moments. The length of the written outline that you prepare is up to you. You decide how complete and detailed the outline should be. If you get your material outlined in less than 90 minutes go back over the material again to assure yourself that the job has been well done. Work at it for the full 90 minutes. Do you have any questions?"

V. Talk Without Notes. "Your study assignment this week is to prepare a 5 to 10 minute talk before a group of your friends, just a

small group, to be given without any notes at all. In fact, you are not going to be permitted even to make any notes while you are studying the material and planning your talk.

It should be important to you to cover as much of the information as possible in the short time allowed for the talk. Don't think about jokes and stories that fill up the time without providing any of the important information about the topic. Are there any questions? You might want to read over the entire material before planning your talk. That is up to you. In any case, spend the full 90 minutes on this assignment."

VI. Written Notes for Talk. "Your study assignment this week is to prepare a ten minute talk on the topic. This is to be a talk before a small group of your friends. You are to make notes to help you remember what to say, but you are not to prepare to read a paper to them. How you say things is not going to be important. Remember, how you say things is not going to be important, but the information that you cover in your talk is important. Are there any questions?

You might want to read the entire material before planning your talk. That is up to you. In any case, spend the entire 90 minutes on this assignment."

VII. Read Programed Test. "Your study condition this week is going to be with programed textbook materials. I have the lesson sheets here for you. You have had some practice with these programed materials. You remember that you fill in the blanks and then check to see if you made the right answers. Be sure to write the answers on the sheets in the spaces provided. Answer every frame in writing. Are there any questions?"

VIII. Write Programed Text. "Your study condition this week is going to be with programed materials. I have the lesson sheets here, and will distribute them to you in just a few moments. You have had some practice with these programed materials. You remember that you fill in the blanks and then check to see if you made the right answers. Today you are not going to write the answers down. You are going to read through the frame, to think of the right word or words to complete the sentence, but not to write anything. Are there any questions?

You probably will not complete all of the program within the 90 minutes. If you should complete it, go back to the beginning and study through some more of it."

IX. Read Teaching Machine. "This week you will be using teaching machines to help you with your studying of the materials. Five of you should have the green Minimax machines and five of you should have the gray Conceptograph machines.

You will remember that when using teaching machines, you fill in the blanks and then check to see if you made the right answers. Today, you are not going to write the answers. You will read the frame, and

then think of the answer that goes in the blank, but you will not write anything. You will check to see that the answer you are thinking about is correct, and go on to the next frame. Are there any questions? You are probably not going to be able to get through the whole program in the next 90 minutes, but get as far as you can."

X. Write Teaching Machine. "This week you will be using teaching machines in your studying. Half of you will be using the green Minima machines and the other half will be using the gray Conceptograph machines. At a session later this spring, each of you will shift to the other machine.

You will recall that when you use these teaching machines, you fill in the blanks and then check to see if you made the right answer. You will read the frame, think of the answer, and then write the answer in the place provided for the answer. Be sure to write an answer for every frame. Are there any questions? You probably will not be able to finish this whole program in the 90 minutes, but study as much of it as you can."

Measurement of Achievement

A twenty minute quiz was administered on each topic two weeks (three weeks at spring vacation time) after each study period. All but one quiz contained three parts; the other had two parts. The internal arrangement varied slightly from quiz to quiz, but on most the first part was an item for open-ended recall, e.g. "Write a paragraph which covers the reasons for buying life insurance." The second part contained completion items, e.g., "A premium which is computed by multiplying the face amount times the probability of dying is called a _____ premium." The last part of the quiz contained multiple choice items, e.g. "The economic principle of life insurance is that: (a) planning ahead is good sense, (b) financial hardship can be avoided by spreading several severe losses among many participants, (c) life insurance is preferable to any other economic source as financial protection for the aged, (d) as time passes the life expectancies of people decrease."

These different item types were used because it was believed that some conditions might result in achievements best measured by one item type whereas other conditions might result in achievements best measured in other ways. The scores from 63 quiz parts were factor analyzed and no factors emerged to substantiate this interaction hypothesis (Table 4). Therefore each part of the quiz was given about equal weight and a single total score was recorded for each 20 minute quiz.

To provide information on retention over a longer period the quizzes for topics studied in the first five weeks were administered a second time in the final week, week #14. On the average, ten weeks had elapsed since study of the material and eight weeks since the previous administration of these quizzes. As before, the full twenty minutes was

taken for each quiz, but the fact that five quizzes were administered back-to-back might have affected the results. The fact that paychecks were available following the last quiz may have countered some loss of interest.

Other Measurement

During the quiz-time immediately following each of the first two study sessions, and during the study-time of the last two sessions, a total of about 220 minutes were available for other testing. As mentioned in the previous paragraph a portion of this time was used for readministering five quizzes. The remainder was used for obtaining an indication of verbal ability, an indication of value-orientation, and an indication of attitude toward study habits and programed material.

Verbal ability was indicated by scores on the Borgatta-Corsini Quick Word Test. This is a 100 item vocabulary test for adults.

The Foe Inventory of Values (FIV) was administered so that we could answer a question about the relationship between personal values and satisfaction and/or success with different study situations. It was thought that there might be an objection to teaching machines by the students with high scores on the aesthetic scale. The scales of the FIV are:

1. Aesthetic
2. Intellectual
3. Material
4. Power
5. Social Contact
6. Religious
7. Prestige
8. Humanitarian

Attitude toward modes of study were examined both before and after the experimental sessions. The questions asked are listed in a separate section in the next chapter.

Miscellaneous Information

The groups all met in the same room in the basement of the University of Nebraska Teachers College. This was a well lighted room, equipped with tables and chairs, (a table per person, if desired), teaching machines, and a supervisor. Graduate students Neil Hartman and Tom Osborne each had five of the groups to supervise. They distributed materials, read directions, kept time, fixed machines, and on occasion, suggested that a student resume his study or looked unhappy if he hadn't. During some of the sessions they observed and recorded the apparent diligence of the students. In the early weeks they read an occasional plea for sincerity and typical study performance.

The teaching machines were of two makes, TMI Minimax and Conceptograph. These were two of the manual, relatively inexpensive, machines being distributed at that time. Some of the machines were not in working order when received, but the manufacturer supplied the necessary replacement parts, and repairs were made. All machines were still in working order at the end of the 14th week. Each machine had been used about four hours per week.

Half of the students working with teaching machines used one make, the other half the other. They alternated makes the next time they used machines. Other than during the orientation session, the students spent only 90 minutes with any one make of machine. Some students complained that the machines were cumbersome.

During the first week orientation session the general purpose of the research was explained to the students. They were told something about programmed instruction, obligated to respond to several short linear programs, introduced to the machines, and given opportunities to use them. They were also given another brief program to be completed at home.

Apparent Study Activity

During four sessions in the later weeks of the experiment the "study hall supervisors" secretly recorded the number of students in the group present who appeared to be studying. Quite a few arbitrary decisions had to be made, of course. The results were combined on various bases as is indicated in Figures 3, 4, and 5.

After 25 minutes of study, the students assigned to read, to plan, and to operate machines, but who were not permitted to write began to lag (Figure 3). The "writers," on the other hand, were noticeably less easily distracted for the remainder of the 90 minute period. At the end there were five percent more of the "writers" studying than "readers." But the surprising thing is that there was no overall difference (across 10 weeks) between achievement scores for these two groups. The readers appeared to learn as much even though they were less tenacious in their study.

Students of programmed materials were obviously more involved in their work than students with conventional materials (Figure 4). Here the apparent studiousness was related to achievement: programmed performances excelling. The means of 51.5 and 49.0 are significantly different, as will be explained in more detail in the next chapter.

In Figure 5, the ten different study situations are presented separately. Students in Situation X appear to be the busiest of all--too busy probably (or busy at the wrong business) because their achievement is much below that of the other three groups of programmed learners. In Situations I - VI the consistency of overt study activity for "writers" over "readers" is apparent. Yet the quiz scores indicated equal achievement.

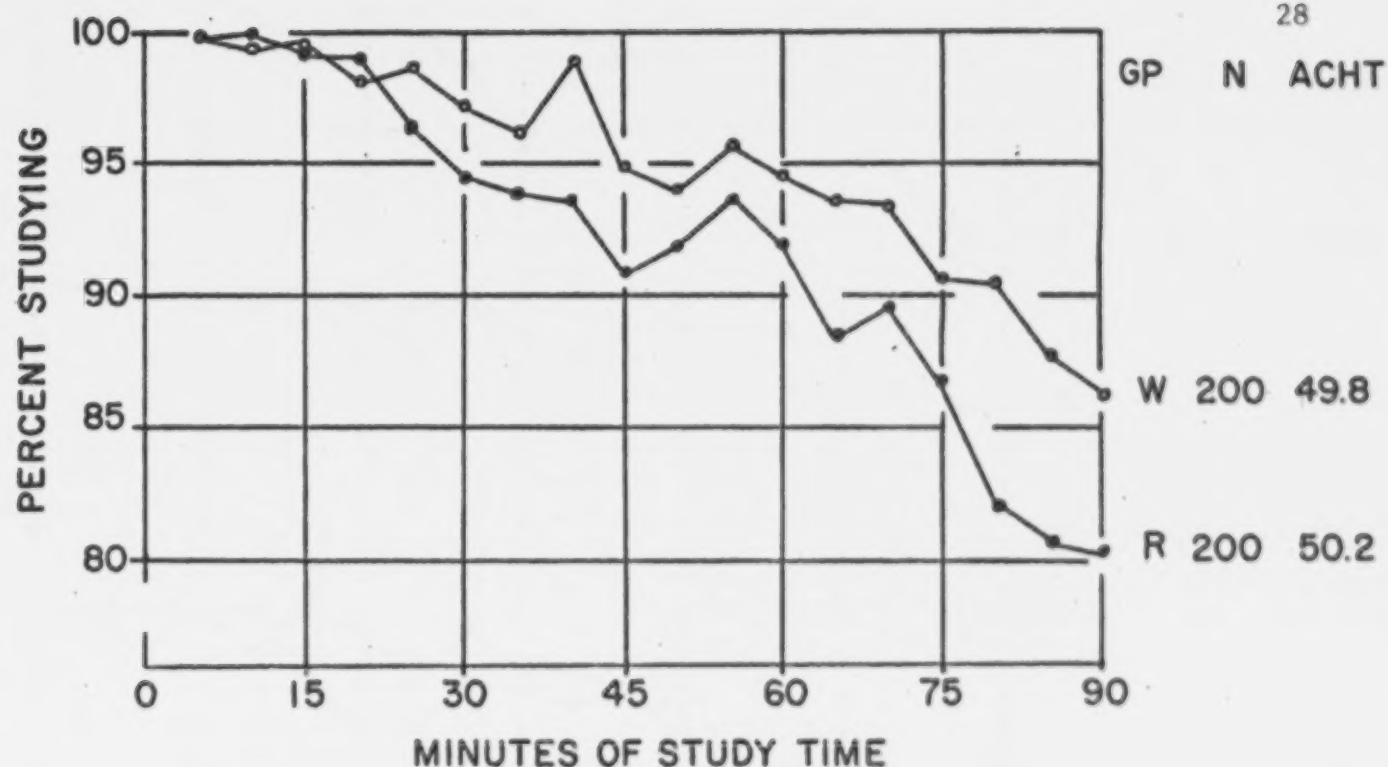


Figure 3. Comparison of the apparent study involvement of students (W) making written responses during study with students (R) studying without making written responses.

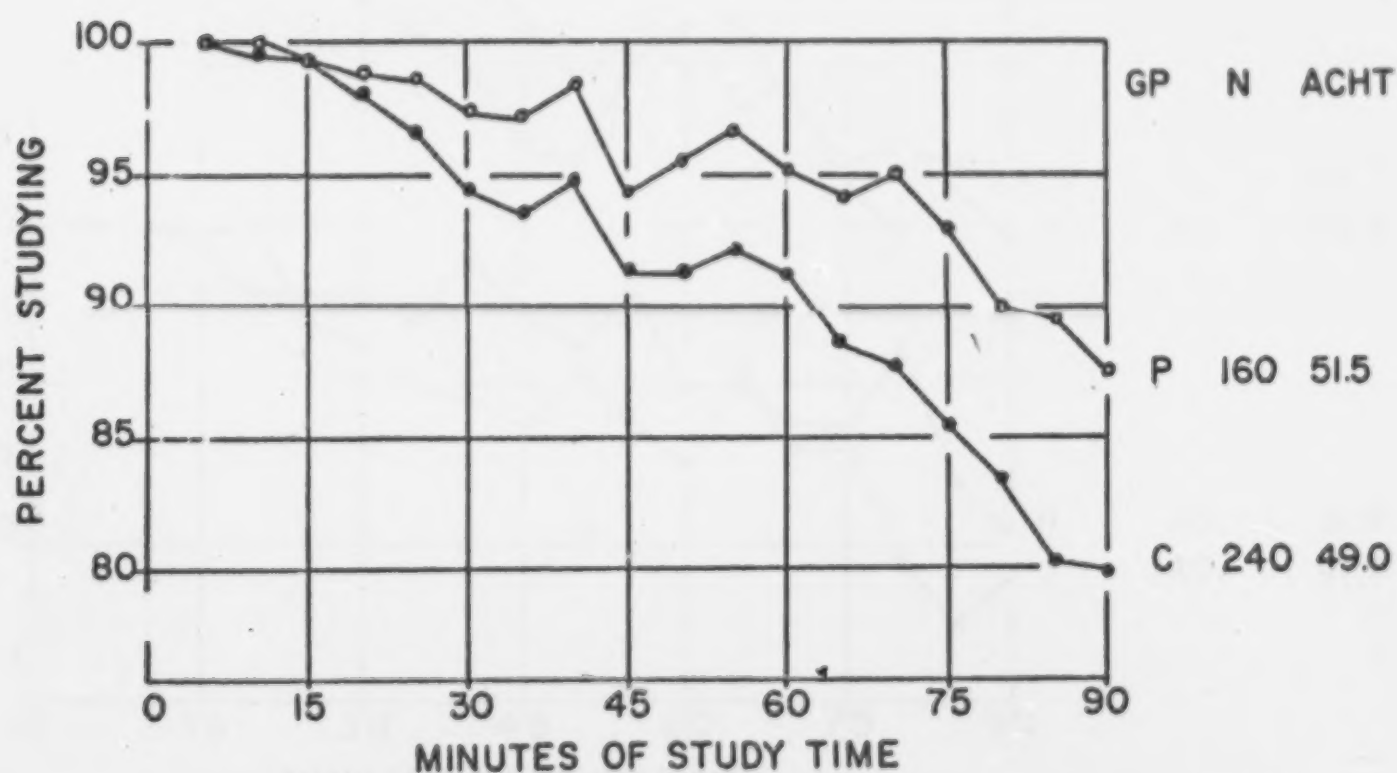


Figure 4. Comparison of the apparent study involvement of 160 students (P) studying programed materials with 240 students (C) studying conventional materials during 90 minute study periods.

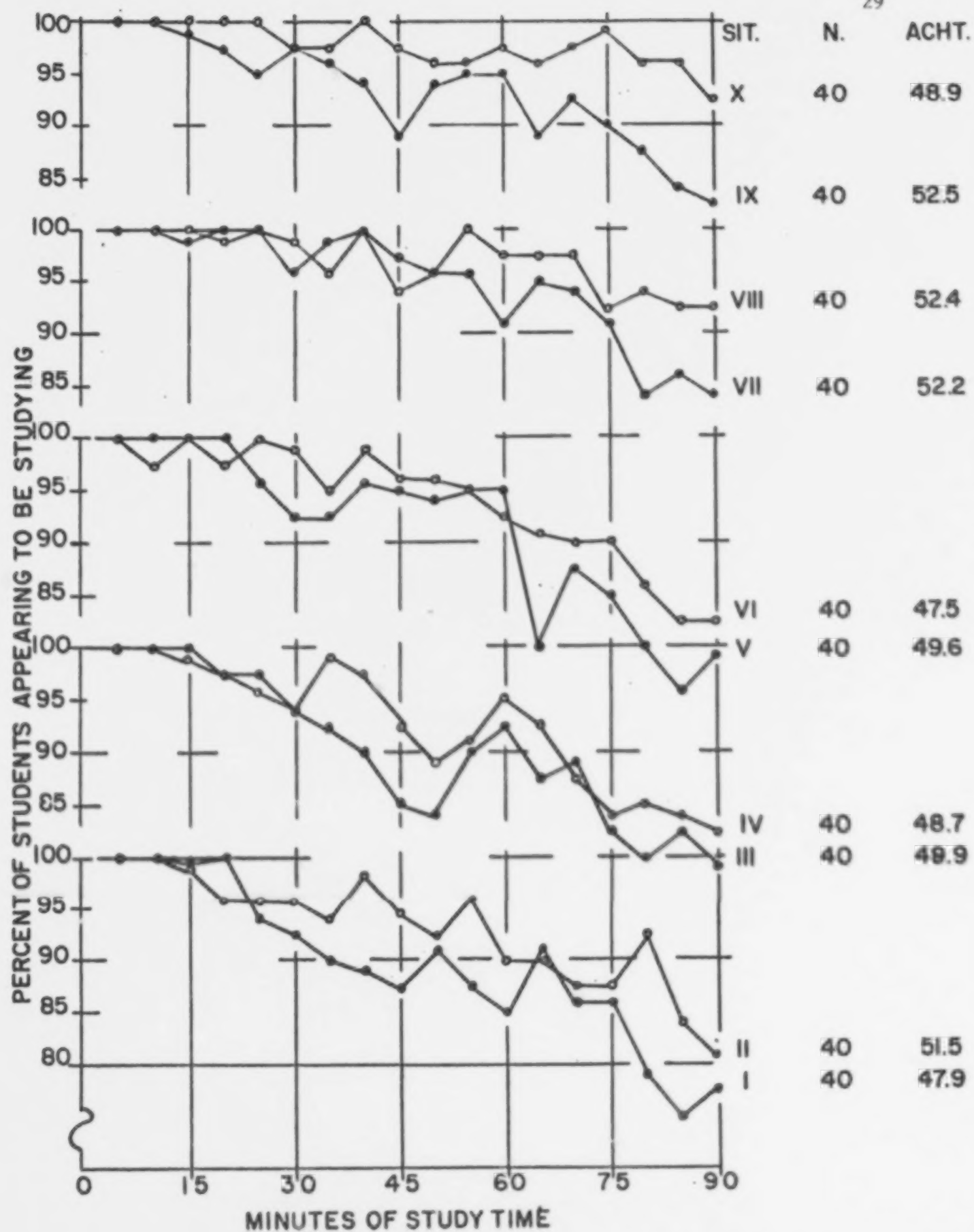


Figure 5. Percent of students appearing to be studying under 10 different study conditions during 90 minute study periods.

Statement of Hypotheses

The ten main hypotheses of this research project and a summary of findings are stated below:

Hypothesis 1: Achievement will increase if students are required to respond in writing during study. In other words, the very act of writing notes, outlines, and answers which otherwise would be made only covertly if at all, will enhance learning. In keeping with this hypothesis it is presumed that the recording of a thought or an association in writing is a step toward assimilation and retention of the subject matter.

In our research project, the mean of all performances involving writing was almost identical to the mean of other performances; but inconsistencies were found. There was an interaction effect between writing and other circumstances of instruction. Although it seemed that it helped to make written notes rather than mental notes, it seemed that it hindered to make written responses with the teaching machines used in this study. Hypothesis 1 was partially supported; under some circumstances writing appeared to facilitate learning.

Hypothesis 2: Achievement will increase if the material is programmed. Where the same students are studying both conventional and programmed materials it is hypothesized that the instruction based on programs will be the more effective. It was recognized that these two types of instruction differ at least a little with respect to content, the response set called for, the amount of response evoked, the type and amount of reinforcement, and the novelty of the situation.

In our study conventional materials were used in six situations; programmed materials, in four. The difference in achievement was statistically significant, favoring the programmed format. The hypothesis was supported.

Hypothesis 3: Achievement is enhanced by the novelty of the instructional situation.

Programmed textbooks and teaching machines were presumed to be novel. To investigate the question of whether or not novelty itself is a determinant of achievement, two additional novel situations, V and VI, were included. It was found that achievement was somewhat better in the six novel situations, but the differences in means was not great enough to be considered substantial support of the hypothesis. It was concluded that greater achievement with programmed instruction was attributable to something other than novelty.

Hypothesis 4: Instructional situations vary as to the amount of overt study activity evoked. It is hypothesized that overt study activity will be more intense where writing is assigned and where programmed materials are used. This hypothesis was confirmed.

Hypothesis 5: The difference in achievement due to instructional conditions can be described better with reference to a multidimensional structure of impacts and personal susceptibilities than with reference to categorical effects on group behavior. In other words, impact analysis, a multidimensional regression model described in this research report, can account for more nonerror variance in achievement than the traditional analysis of variance model.

As is usually the case in a repeated measurements design, the main effect for persons accounted for much more of the variance than the situation "categorical" effects. The impact analysis did not appear to reduce the error variance significantly below the analysis of variance model involving person means. The hypothesis was not confirmed.

Hypothesis 6: The ten instructional situations used in this study can each be assigned a point on a unidimensional scale which runs from situations of low overt activity yielding relatively less achievement to situations of high overt activity yielding better achievement. To support this hypothesis it was necessary to demonstrate that activity and achievement do differ from situation to situation, that there is a correlation between activity and achievement, and that access to more than a single dimension of activity does not appreciably reduce the error variance.

There were significant differences in levels of achievement for the ten situations. Level of achievement was positively correlated with anticipated study activity and with observed study activity, but to a lower degree than expected. Other bases for high achievement were sought.

Hypothesis 7: Some types of quiz items are biased toward certain methods of instruction so that students will appear to have learned more or less than they did, depending on whether the item type and instruction match or not. That is to say, there is an analysis of variance interaction between the instructional method and the evaluative method. In particular it is supposed that students taught by linear programs will be favored on completion-type items whereas students taught in more holistic fashion would do better on essay-type items and whereas students learning with branching programmed material would be favored when evaluated with multiple choice items.

With the design explained in this chapter it was deemed impractical to try to test the interaction of item type and instructional condition by analysis of variance. On the grounds that this same source of variance, if operating, would determine factor analysis factors, factor analysis was used. No major factors were interpreted as having been determined by item types. It was concluded that if there were any item-type bias in these data at all, it could be considered negligible. The hypothesis was not supported.

Hypothesis 8: Long term retention of acquired learning is a function of the instructional method. It is hypothesized that writing, that

using programed materials, and that devising novel circumstances will increase the amount of material retained by the student. Whereas earlier hypotheses were devoted to achievement measured by tests of retention two weeks after acquisition, this hypothesis is devoted to achievement measured with the same tests ten weeks later.

In this study, after the 10 weeks it was found that much of what had been acquired had been forgotten and that the loss often was greatest where the acquisition had been greatest. Regression effect was considered. Greater losses however, were not consistently associated with any particular instructional method. The hypothesis was not supported.

Hypothesis 9: After participation in the study, students will be of the opinion that topics which are programed are both more enjoyable and easier to learn. Some support for this hypothesis was found, although in general the participants were not as favorably disposed toward programed instruction at the end of their work than they had been at the outset.

Hypothesis 10: The willingness of students to accept programed material and the amount they learn in using it are related to the personal values of the students as manifested in responses to a paper and pencil inventory of values. No support was found for this hypothesis.

CHAPTER III

UNIVARIATE ANALYSIS OF ACHIEVEMENT

The analysis of achievement was based primarily on data from the ten quizzes which were administered two weeks after the study sessions. Each of these quizzes contained different item types. A subscore was obtained from items of a given type within a quiz. Each of the quizzes yielded three subscores except the English Money quiz which yielded two. Thus, a total of 29 subscores were available from one administration of the 10 quizzes. When the quizzes for the first five topics were readministered at the end of the study, 14 more subscores were available for analysis.

The pretest, not to be confused with the quizzes of the previous paragraph, contained completion items on all ten topics. Ten pretest scores were available, then, from the Week 1 administration, and ten more from the posttest session. When these 20 variables were assembled with the 29 and 14 previously mentioned, we had an array of 63 achievement scores per student. The product moment coefficients of correlation were calculated for all pairs of variables.

Item-Type Bias

To satisfy ourselves that item types would not bias the measurement of achievement, we studied the variance of scores associated with item types. We reasoned that if there were a contingency between instructional method and quiz type, some factor analysis factors would be identified by clusterings of quizzes of a given item type.

A centroid factor analysis was made. Diagonal entries, originally largest zero order coefficients, were re-estimated as each factor was extracted. Six factors had roots larger than unity. The configuration was rotated to varimax criterion. From the complete factor matrix, the abridged matrix shown in Table 3 was made. Here, the 39 row variables represent the 39 independent achievement variables, each based on a single item type. No posttest or retest variables are included in Table 3, although they were included in the rotated factor matrix from which Table 3 was made. The factor loadings from the largest ten factors are shown, rounded to one decimal place, although probably no more than five or six factors were prominent enough to deserve interpretation. The essay portion of the Insurance quiz, for example, had a loading of around .60 on the sixth factor. The last column of Table 3 indicates communalities and the last row indicates factor variances. It was realized that the factor structure was influenced by the inclusion of many retest variables.

Table 3

Factor Loadings and Communalities of Variables Sorted as to Item Type

	FACTOR										² h
	I	II	III	IV	V	VI	VII	VIII	IX	X	
Essay Items											
Insurance	0	2	1	2	1	6	2	1	+1	+1	46
Poetry	5	2	-1	0	0	0	2	2	0	+2	43
Nebr Govt	1	1	0	2	2	1	3	1	0	+3	34
Leaves	2	1	0	3	0	2	4	1	0	+1	37
Art	2	1	3	2	2	0	2	0	+	0	29
Ferment'n	-1	-1	1	3	4	0	0	0	+5	+1	54
Performance Items											
Logic C	1	2	2	4	1	-1	1	3	-1	+1	42
Logic B	-1	-1	1	6	0	1	-1	0	-2	+1	39
Leaves	1	1	1	-1	1	1	1	0	+5	0	26
Dewey Dec	1	7	1	0	1	1	0	1	0	0	58
Completion Items											
Insurance	2	0	2	1	1	6	1	0	0	+1	43
Dewey Dec	1	7	1	1	0	2	0	0	0	+2	66
Eng Money	1	4	7	2	2	1	0	1	+2	+2	80
Poetry	8	1	1	2	1	0	2	0	+2	+1	82
Logic	3	1	1	4	4	1	1	1	-1	+1	56
Nebr Govt	2	1	2	0	1	4	4	3	+4	0	67
Leaves	3	2	3	3	4	0	3	-1	+1	+1	57
Ger Money	1	3	3	2	2	1	0	2	+4	+3	68
Art	1	1	2	4	3	3	0	3	+3	-1	49
Ferment'n	3	2	4	2	4	1	2	3	+2	+2	68
Pretest Completion Items											
Insurance	-2	0	3	-2	0	3	0	0	0	+1	26
Dewey Dec	2	2	3	-2	0	1	4	2	+3	+1	51
Eng Money	1	1	-1	2	-1	2	1	2	+3	-2	33
Poetry	6	1	0	-1	-2	0	1	0	+1	-1	50
Logic	-1	0	1	1	1	0	0	0	0	-1	24
Nebr Govt	-2	1	2	-2	2	4	3	1	+1	+2	45
Leaves	-2	2	2	0	4	3	0	1	+3	-2	43
Ger Money	0	1	2	1	0	1	-1	0	+1	+5	34
Art	0	-1	4	0	1	-1	0	3	+2	-2	35
Ferment'n	1	0	1	-1	5	3	0	-1	+1	+1	42
Matching Items											
Ger Money	2	1	1	0	2	1	1	1	0	+6	51
Multiple-Choice Items											
Insurance	2	0	2	4	2	3	0	4	0	0	69
Dewey Dec	2	3	1	0	-1	-1	3	5	0	-1	53
Eng Money	3	2	6	1	2	1	1	2	-1	+2	70
Poetry	6	2	0	2	3	0	0	0	+1	+1	61
Nebr Govt	2	1	1	1	0	3	6	2	+3	0	70
Ger Money	1	2	2	3	2	0	0	5	+3	+3	64
Art	3	1	2	3	3	3	0	1	+3	-2	51
Ferment'n	1	0	2	2	5	1	0	4	+1	+2	60
Factor Variance											
	48	40	39	37	33	26	26	24	23	23	

Factor VIII seemed to be the nearest to a factor which might be interpreted as an item-type factor. The largest six loadings were associated with multiple-choice subscores. However, none of the next six highest loadings were with multiple-choice items and this factor was one of the ones that usually would be labeled an error factor. Furthermore, none of the other factors could be considered an item-type factor. Therefore it was concluded that no important component of achievement could be traced to item-type bias.

If the prominent factor loadings of Table 3 are rearranged with reference to topic, as they have been in Table 4, it is obvious that topic was the basis for the factorial structure of this set of variables.

As far as item types are concerned, we did not necessarily use unbiased tests. We merely satisfied ourselves that there was not a sizable interaction between instructional condition and item type. If any substantial interaction of this kind existed, there would have been a substantial factorial variance associated with the bias. One or more factors would have been identified in terms of item type. Finding such factors would not have assured us of bias due to instructional method, but failure to find it satisfied us that what bias there may be is negligible.

This preliminary step to analysis was not a critical one. We had counter balanced the design so that all persons participated in all instructional situations and answered quiz items of several types after studying each topic. The check on item bias assured us that we were not missing an important source of variance. With no substantial bias due to item types, there seemed to be no reason to use the quiz subscores further. They were combined by simple addition to make a single achievement score for each topic.

Pretest to Posttest Gains

As indicated in the previous chapter, four completion type items for each topic were administered before and after the ten learning sessions. Though these tests were brief, a statistically significant gain appeared with study of each topic. The mean items right and "t" values are shown in Table 5. A correlated t-test was employed. All ten differences were found to be significant at the .001 level. There was no doubt about whether or not some achievement had occurred.

Quiz Reliabilities

The 10 composite quizzes did not include the items of the pretest and posttest. Reliabilities for these 10 quizzes were estimated on the basis of the communalities and intercorrelations of their parts. For the first five quizzes the correlation with retest scores was also considered. Estimates of reliability coefficients are given in Table 4.

Table 4

Characteristics for 63 Achievement Quizzes

Testing Time, Test Retest Correlation,
Loadings on the Five Largest Factors, and Communality,
and Estimated Reliability Coefficients
of the 10 Composite Quizzes

	Min.	r_{12}	I	II	III	IV	V	h^2	r_{xx}
Insurance									80
Pretest	2	34	-2	0	3	-2	0	26	
Posttest			0	1	1	1	2	42	
Essay	5	35	0	2	1	2	1	46	
Retest			2	2	01	2	0	30	
Compl	5	33	2	0	2	1	1	43	
Retest			0	0	4	4	3	62	
Mul-Ch	10	64	2	0	2	4	2	69	
Retest			1	2	0	2	1	45	
Dewey Decimal									80
Pretest	2	40	2	2	3	-2	0	51	
Posttest			2	5	1	2	1	48	
Compl	5	66	1	7	1	1	0	66	
Retest			2	7	2	0	1	73	
Mul-Ch	5	55	2	3	1	0	-1	53	
Retest			2	3	3	1	0	54	
Class'm	10	60	1	7	1	0	1	58	
Retest			3	7	1	0	1	66	
English Money									85
Pretest	2	17	1	1	0	2	-1	33	
Posttest			1	2	5	2	1	42	
Compl	10	76	1	4	7	2	2	80	
Retest			2	3	5	4	2	80	
Mul-Ch	10	63	3	2	6	1	2	70	
Retest			2	3	6	2	1	58	
Poetry									
Pretest	2	46	6	1	0	-1	-2	50	
Posttest			6	2	2	1	2	58	
Essay	5	44	5	2	-1	0	0	43	
Retest			3	2	-1	-1	0	47	
Compl	7	82	8	1	1	2	1	82	
Retest			8	1	2	1	1	78	
Mul-Ch	7	59	6	2	0	2	3	61	
Retest			6	2	3	0	2	62	

(Table continued on next page)

Table 4 continued

	Min.	r_{12}	I	II	III	IV	V	h^2	r_{xx}
Logic									65
Pretest	2	18	-1	0	1	1	1	24	
Posttest			1	1	1	7	1	64	
Compl	5	62	3	1	1	4	4	56	
Retest			2	2	1	4	4	46	
Applic I	7	28	-1	-1	1	6	0	39	
Retest			1	1	1	4	-1	40	
Applic II	5	40	1	2	2	4	1	42	
Retest			0	0	-1	6	0	57	
Nebraska Government									80
Pretest	2	33	-2	1	2	-2	2	44	
Posttest			2	1	0	0	2	54	
Essay	5		1	1	0	2	2	34	
Compl	5		2	1	2	0	1	67	
Mul-Ch	10		2	1	1	1	0	70	
Leaves									65
Pretest	2	41	-2	2	2	0	4	43	
Posttest			1	2	2	1	7	66	
Essay	5		2	1	0	3	0	37	
Compl	10		3	2	3	3	4	57	
Applic	5		1	1	1	-1	1	26	
German Money									85
Pretest	2	30	0	1	2	1	0	34	
Posttest			0	3	2	2	2	56	
Compl	7		1	3	3	2	2	68	
Match	5		2	1	1	0	2	51	
Mul-Ch	7		1	2	2	3	2	64	
Art Logic									70
Pretest	2	25	0	1	4	0	1	35	
Posttest			2	-1	5	1	1	49	
Essay	5		2	1	3	2	2	29	
Compl	8		1	1	2	4	3	49	
Mul-Ch	7		3	1	2	3	3	51	
Fermentation									85
Pretest	2	37	1	0	1	-1	5	42	
Posttest			4	1	3	2	5	67	
Essay	5		-1	-1	1	3	4	54	
Compl	7		3	2	4	2	4	68	
Mul-Ch	7		1	0	2	2	5	60	

NOTE--Decimal points omitted from correlation coefficients, factor loadings, communalities and reliability estimates. Loadings rounded to nearest .10.

Table 5

Topic Means on Pretest and Posttest and the
"t" Values for the Difference

	$\bar{X}_1$	$\bar{X}_2$	"t"
Insurance	1.09	2.93	15.29
Dewey Decimal	1.68	3.80	6.01
English Money	.83	1.36	6.70
Poetry	1.14	2.85	12.14
Logic	.23	1.16	7.98
Government	1.53	2.44	7.54
Leaves	.80	2.00	8.85
German Money	.50	1.69	14.21
Art	1.59	2.12	4.77
Fermentation	.36	2.65	15.48
Total	9.77	23.00	23.54

The scores from each of the ten quizzes were converted into non-normalized T scores. This conversion equated levels and variabilities of achievement on all topics, whether or not any learning occurred and whether or not the measurement of achievement was equally free of error on the different topics. The scores were then combined, across topics, to fit into the analysis which is presented in this chapter. The mean achievement for each situation then was based on 10 quiz scores from each of the 10 topics. The means and standard deviations for all levels of all experimental variables are shown in Table 6.

In column B of Table 6 there are indications of how busy the students appeared to be, on the average. Each percent is based on observations taken every five minutes in the study room. The total of 93% means that 9.3 of the 10 students in the study room appeared to be engaged in study, overtly or covertly, on the average. This total figure was based on observations made during 40 of the 100 study sessions.

The Expected Order of Means

The study situations were originally numbered in pairs from I, a passive situation, through V to X, a teaching machine activity. The odd numbered situations did not involve writing whereas the even numbered situations did. It was believed that the achievement in the even numbered situations would exceed that of the odd.

The pairs were ordered in the degree to which it was believed the students would be personally involved in study. Taking of notes was thought to be least compelling, using teaching machine most compelling. The preparation of a talk was not restrictive, but it was predicted that the student would become more immersed in the assignment in situations V and VI than in situations I to IV and less than in situations VII - X.

Overt study activity is not the same thing as personal involvement, of course, but observations of busyness did conform fairly well to the expected order of involvement. On the basis of clandestine observations of busyness during four of the sessions, apparent involvement in study increased in the following order: Pairs I-II, III-IV, V-VI, IX-X, VII-VIII. (See Figures 4-6 and Table 6). Situations II and IX were the only ones which were quite out of place in the actual order, both generating middling amounts of overt study activity. Within each of the pairs the requirement for writing increased the amount of apparent study activity.

The Observed Order of Means

Achievement in the ten situations is indicated in Table 6 and Figure 6. The order of means for pairs is V-VI, (I-II and III-IV), IX-X, VII-VIII. Except for our anticipation of programed materials to

Table 6
Means and Standard Deviation
of Achievement Scores
and Average Busyness Scores

		N	X	s	B
Writing		500	50.2	9.9	95%
	No writing	500	49.8	10.2	92
Conventional		600	49.0	10.0	92
	Programed	400	51.5	9.9	96
Traditional		400	49.2	10.1	91
	Novel	600	50.7	9.9	95
Situation	I	100	48.2	10.0	89
	II	100	50.2	9.6	93
	III	100	49.9	10.4	90
	IV	100	48.4	10.4	93
	V	100	48.7	10.4	91
	VI	100	47.6	9.0	94
	VII	100	52.0	10.1	95
	VIII	100	53.1	9.1	97
	IX	100	52.5	9.6	93
	X	100	49.5	10.5	98
Group	A	100	52.4	8.7	95
	B	100	47.1	11.2	94
	C	100	49.4	9.3	94
	D	100	50.9	9.6	95
	E	100	49.5	9.1	94
	F	100	47.7	10.7	88
	G	100	52.5	10.8	95
	H	100	47.5	8.9	90
	I	100	50.9	10.1	96
	J	100	52.1	9.9	92
Total		1000	50.0	10.0	93%

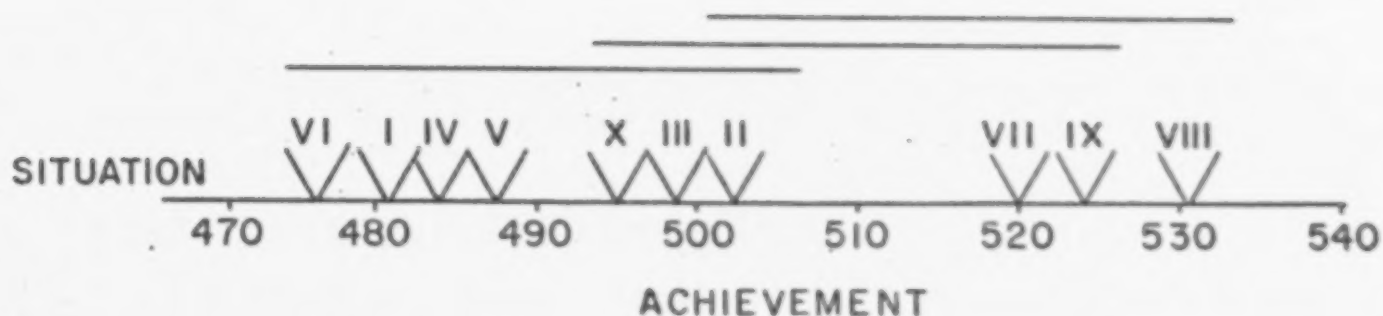


Figure 6. Unidimensional Activity-Achievement scale with means of 10 situations.

excel, our expectations were free of clairvoyance. In only 2 of the 5 pairs did writing during study bring about superior quiz performances.

The correlation between anticipated achievement and observed achievement was not calculated because we had not predicted a complete order of 10. We did not commit ourselves in advance as to whether Situation II or III would be more productive, just that they both would be more productive than Situation I. We will note here, however, that the rank order correlation coefficients between the situation numerals and achievement means and also between the achievement and busyness means are about +.5. The relationship between study activity and achievement was not as high as we would have liked it for the support of Hypothesis 6. Furthermore, what positive relationship there was could almost entirely be attributed to the superiority of programed materials.

Of the three major comparisons based on writing, programed materials, and novelty, only the second proved to be a critical variable. The difference between means of 49.0 and 51.5 was statistically significant at the .01 level. The programs clearly facilitated achievement. Whether the gain is due to the greater activity, due to the difference in content or due to the immediate reinforcement we are not able to say. By including Situations V and VI we have some reason to believe that the difference is not due to the novelty of the methods. The novel instructional situations did not yield achievement significantly different from the traditional situations.

The mean for the 500 scores obtained when study involved writing was compared to the mean for a like number from study without writing.

The means turned out to be almost identical. As can be seen in Table 6, the means for these two groups differ by about .04 of a standard deviation. If no distinction is made as to what the students are to write, requiring them to write while studying apparently does not improve learning.

The comparisons cited in the two previous paragraphs were not made until a complete analysis of variance was obtained. That analysis is to be reported next.

Analysis of Variance of Achievement Scores

In Figure 2, the experimental design is represented as a $2 \times 5 \times 10$ model, representing Mode, Condition, and Topic. Groups of persons are repeated in Latin Square fashion. Let us call this representation, Model A. In a straight-forward three way analysis of Model A scores, variance involving persons would be included with error variance.

The experimental design we used can also be represented by a similar $2 \times 5 \times 10$ structure showing Mode, Condition, and Groups as main effects. In Figure 7 this second model is called Model B. Topics occur as a Latin Square variable. In a third model shown in Figure 8 as Model C, the design is 10×10 with Topic and Groups as the main effects, and situations buried in the Latin Square arrangement.

All three models are relevant. Different sources of variance are obtained by each model. Therefore, some of the variance is confounded. Variance of the main effects: Groups (or Persons), Conditions, Modes, and Topics, is not confounded; but variance of all of their interactions, except Condition by Mode, is partially confounded.

In Table 7 the main effects and interactions are presented for this analysis. All confounded interactions are presented under the heading of error. The error can be parcelled out three different ways. With Model A, error includes Situation by Topic interaction, and variance within Situation by Topic cells. Since these latter two sources include variance for persons, to treat Persons as a main effect we must subtract the associated sum of squares and degrees of freedom to obtain the total error terms.

With Model B, error includes Situation by Person interaction, but does not include that part of this interaction identified with Topics. With Model C, error includes Topic by Person interaction, and Situation deviation squares and degrees of freedom must be removed to obtain the total error term.

The reader is reminded that the "error" interactions are confounded. No one of these models is the right one, and none is wrong. No source of interaction within the error portion can be considered

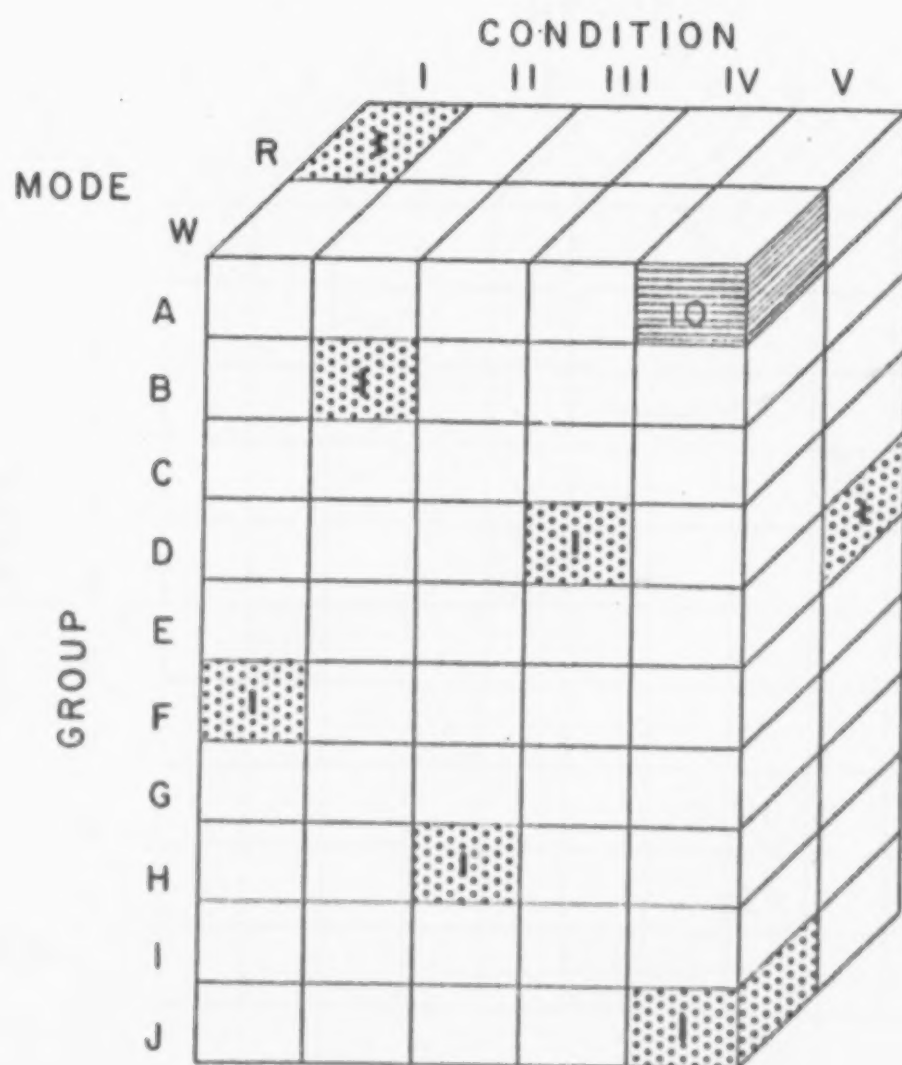


Figure 7. Model B, a structural representation of the experimental design.

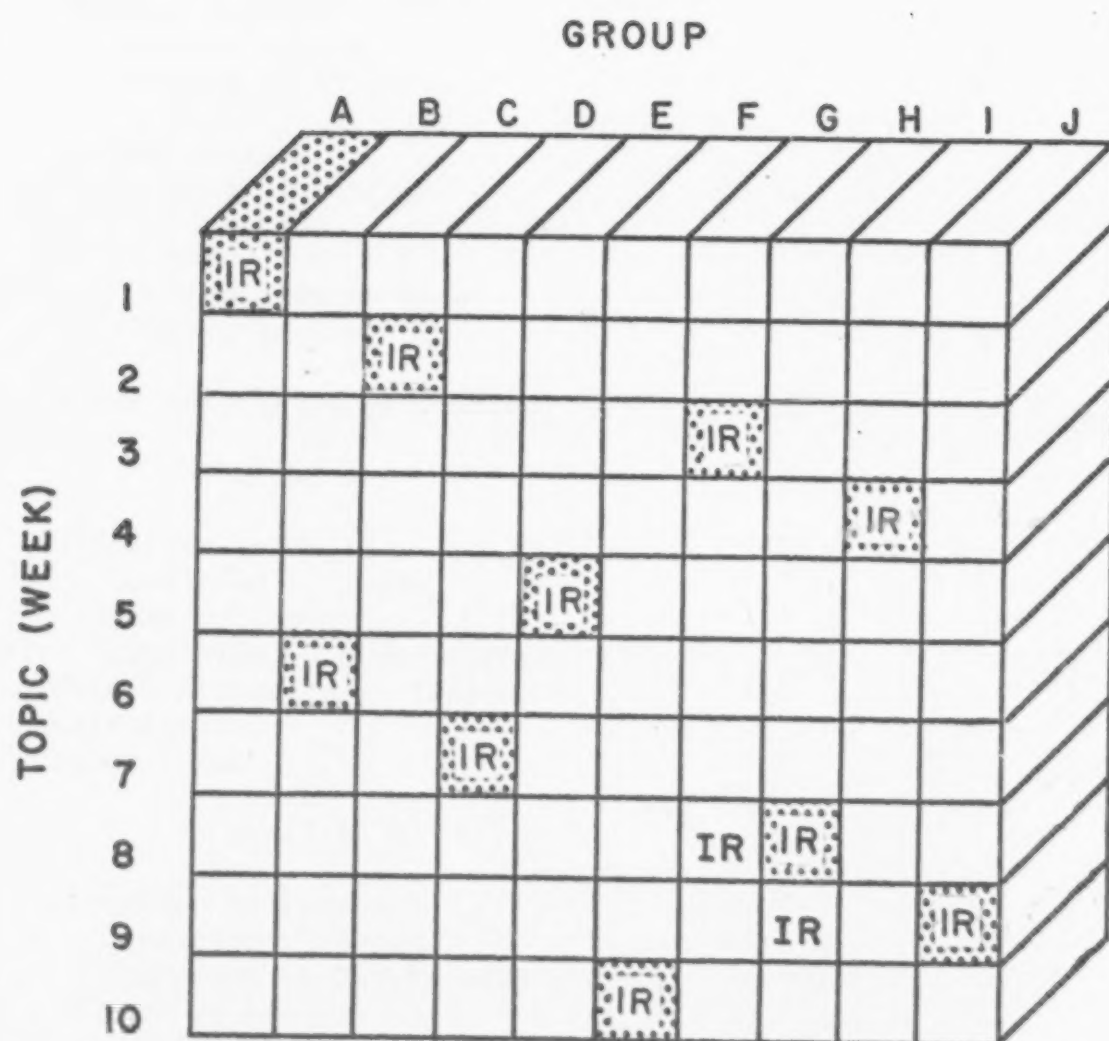


Figure 8. Model C, a structural representation of the experimental design.

Table 7

Main Effects and Interactions

Source of Variance		Apparent Degrees of Freedom	
P	Between Persons		99
G	Between Groups	9	
P	Persons within Groups	90	
	Within Persons		18
S	Between Situations	9	
C	Condition	4	
M	Mode	1	
CM	Condition by Mode	4	
T	Between Topics	9	
E	Error (partially confounded)		<u>882</u>
Model A			
ST	Situation by Topic		81
CT	Condition by Topic	36	
MT	Mode by Topic	9	
CMT	Condition by Mode by Topic	36	
	Within Situation by Topic		900
	Less Person		<u>-99</u>
	Total Error		882
Model B			
SP	Situation by Person		891
CG	Condition by Group	36	
Cp	Condition by Person within Group	360	
MG	Mode by group	9	
Mp	Mode by Person within group	90	
CMG	Condition by Mode by group	36	
CMp	Condition by Mode by Person within group	360	
	Less Topic		<u>-9</u>
	Total Error		882
Model C			
TP	Topic by Person		891
TG	Topic by group	81	
TP	Topic by Persons within group	810	
	Less Situation		<u>-9</u>
	Total Error		882
Total			999

independent of the interactions provided by a different Model. This should not be seen by the reader as a penalty for using repeated measures and a Latin square design, but as a burden one accepts when he considers the many and various effects on a dependent variable such as achievement. In our experiment certain of these effects are only partially controlled. In many studies they are not controlled at all. In experimentation with learning some of these effects cannot be more than partially controlled. As with the passing of virginity, there is no possible way of knowing what might have been, under other circumstances.

Components of Variance

Persons, Groups, and Topics were considered random classifications (Green and Tukey, 1960). Mode (writing-no writing) clearly was a fixed classification. Considerable thought was given to the assignment of the Conditions variable. This variable was to be represented by different levels of activity, so in a sense it was a continuum characteristic. Such characteristics are neither completely fixed nor random classifications.

Obviously the instructional situations chosen were not randomly selected. Some attention was to be given to the different conditions, their order, and their differences, if they were found to be statistically significant; so the random tag seemed inappropriate. In spite of the interest in generalizing to other activity conditions, Conditions was listed as a fixed classification.

Following Green and Tukey, the components of variance of all sources are shown in Table 8.

The Mean Squares

In Table 9 the sums of squares and mean squares are presented. The variance of Topics is zero because scores for each topic were transformed to T scores.

As explained in a previous section, with a Latin square design some main effect variance is included in an interaction mean square. To remove these main effects for the 3 models, the negative entries appear. The subtraction provides a remainder which has been labeled, "total error".

Although the error term interactions are confounded, some ideas may be gotten from the mean squares of these interactions. In Model A the apparent interaction includes a large variance associated with persons ($MS = 501.17$). Since the Condition by Topic, Mode by Topic, and Condition by Mode by Topic interaction mean squares are about the same size, and since their excess over error ($MS = 54.65$) can be attributed to individual differences, these interactions can be presumed to be negligible.

Table 8

Components of Variance for Analysis of Variance of Achievement Scores

Source	Components of Variance	Type of Variable
Between Persons		
Groups	$G + TG + p + Tp + E$	Random
Persons in groups	$p + Tp + E$	Random
Within Persons		
Situation		
Condition	$C + CT + CG + Cp + CTG + CTp + E$	Fixed
Mode	$M + BT + MG + Mp + MTG + MTp + E$	Fixed
CM	$CM + CMT + CMG + CMp + CMTG + CMTp + E$	
Topic		Random
Error		
SP		
CG	$CG + CTG + Cp + CTp + E$	
Cp	$Cp + CTp + E$	
MG	$MG + MTG + Mp + MTp + E$	
Mp	$Mp + MTp + E$	
CMG	$CMG + CMTG + CMp + CMTp + E$	
CMp	$CMp + CMTp + E$	
ST		
CT	$CT + CTG + CTp + E$	
MT	$MT + MTG + MTp + E$	
CMT	$CMT + CMTG + CMTp + E$	
TP		
TG	$TG + Tp + E$	
Tp	$Tp + E$	

Table 9

Complete Analysis of Variance of Achievement Scores

Source of Variance	Degrees of Freedom	Sums of Squares		Mean Square
Between Persons	99		49616.200	501.17
Groups	9	3805.380		422.82
Persons in groups	90	45810.820		509.01
Within Persons	18		3357.500	
Situation	9	3357.500		373.06
Condition	4	2480.470		620.12
Mode	1	59.536		59.54
CM	4	817.494		204.37
Topic	9	0		0
Error	882		48201.300	54.65
ST	81	9909.100		122.33
CT	36	3907.030		108.53
MT	9	1252.464		139.16
CMT	36	4749.606		131.93
Within	900	87908.400		97.68
Less P	-99	-49616.200		501.17
SP	891	48201.300		
CG	36	3131.050		86.97
Cp	360	31636.780		87.88
MG	9	1539.084		171.01
Mp	90	4661.580		51.80
CMG	36	1433.586		39.82
CMp	360	5799.220		16.11
Less T	-9	0		
TP	891	51558.800		57.87
TG	81	9461.220		116.81
Tp	810	42097.580		51.97
Less S	-9	-3357.500		373.06
Total	999		101175.000	

In Model B we have six interactions that could not have been tested at all, even if there had been no confounding: Condition by Persons within Group, Condition by Topic by Persons within Group, Mode by Persons within Group, Mode by Topic by Persons within Group, Condition by Mode by Persons within Group, and Condition by Mode by Topic by Persons within Group. It does appear that Condition by Group and Condition by Topic by Group are negligible. The interaction terms Mode by Group, Mode by Topic by Group, Condition by Mode by Group, and Condition by Mode by Topic by Group are not sufficiently sizable to make us wish that the confounding did not obstruct a test of significance.

In Model C, even without confounding, we would not be persuaded that an important Topic by Group interaction exists. We are completely unable to consider any interaction between Topics and Persons. It is reasonable to presume that they exist. Our speculative conclusion is that we do not have any interactions in the total error term of such magnitude as to require consideration when we interpret our main effects.

According to the rules of aggregating as set forth by Green and Tukey, all the terms listed in Table 8 as error could not be combined; but in view of the confounding and because none of them appeared to be close to statistical significance, they were pooled to make one error term. The Group mean square (MS) was not aggregated with Persons MS because Topic by Group MS could not be aggregated with Topic by Persons MS.

Error Terms

An inspection of Table 8 will reveal that none of the five main effects nor the single unconfounded interaction can be tested directly. For none is there a single denominator mean square containing all the components of variance needed for an exact test. Approximate tests with composite error terms are possible, although some of the components needed here for composites are confounded. Another alternative would be to assume that certain components are zero. Still another alternative technique for obtaining error items is available. A conservative F-test can be made by testing the critical effect against the non-critical component having the largest mean square. Although this approximate test would often lead to a Type II error, it almost certainly does not do so with these data.

The mean square for error terms, using this conservative estimate of error, are shown in the second column of Table 10. The error term for Condition, for example, takes the mean square of 88, that for the Condition by Persons interaction.

Independence of Observations

When scores are obtained from the same persons over successive classes, as they were in this study, the F-test assumption of

independences of observations is violated. Lack of independence alters the distribution of F ratios with the given degrees of freedom by shifting it to the right, increasing the likelihood of a Type I error. The F ratio obtained can be evaluated by acknowledging the reduction in degrees of freedom. The appropriate degrees of freedom are $(k-1)\theta$ and $a(k-1)\theta$ where k is the number of levels of the factor being tested, a is a constant, and θ is an index of the homogeneity of variances and covariances. This θ has a maximum value of 1 when variances and covariances are equal and a minimum value of $1/(k-1)$. A conservative test of the hypothesis can be made by considering θ to be a minimum which amounts to dividing the apparent degrees of freedom by $k-1$ (Greenhouse and Geisser, 1959).

In our case, persons provide repeated measures over Mode, Condition and Topic, but independent measures for Groups. To make the conservative test, Mode degrees of freedom are not altered because they are merely divided by $k-1 = 2-1 = 1$. The degrees of freedom for Condition and Topic would be divided by 4 and 9 respectively to make the conservative tests. The degrees of freedom for interactions involving both Condition and Topic would be divided by 36. Error term degrees of freedom must be diminished as well as the numerator term. Such reduction is unnecessary, of course, if the observed F ratios are not greater than table values for the apparent degrees of freedom.

Table 10

Summary of Analysis of Variance of Achievement Scores

Source	Mean Square	Denominator	F	D.F.	Level of Significance
Condition	620.12	88 (Cp)	7.0	1,90	<.01
Mode	59.54	171 MC)	<1		
Topic	0				
Group	422.82	509 (p)	<1		
person	509.01	52 (Tp)	9.8	90,810	<.01
CM	204.37	55 (E)	3.7	1,220	<.06
Error	54.65				

Comparison of Pairs of Study Situations

Among the 10 study conditions there are 45 pairs of situations which can be compared by Tukey's W procedure (Winer, 1962, p. 207). Using the error mean square of 54.65 with 882 degrees of freedom we found a critical difference of 3.3 points on the activity-achievement scale.

Separations between 14 pairs of means exceed or approximate this difference. These separations, therefore, can be treated as significant at about the .95 level of confidence.

On Figure 6, three horizontal lines have been extended above the achievement scale. Any pair of situations spanned by a common line should not be regarded as significantly different.

Results of the Analysis of Variance of Achievement Scores

When all the estimates, approximations, and conservative leanings were effected, the final analysis appeared as it is presented in Table 10. Mode and Group effects are seen as negligible. Condition and Person effects are seen to be prominent. The Condition by Mode interaction is of sufficient magnitude to justify further inquiry.

We learned from this analysis of variance that the different instructional situations in our study do yield different levels of achievement. The learning was more effective in situations wherein programed materials were employed. Making written responses to the programs apparently facilitated learning when programed texts were used, but apparently became debilitating when the students were actively operating the machines.

But there were differences in achievement not explained by differences in the materials used. Three clusters of situations were found. The situations yielding highest achievement involved the programed learning. The second cluster included a teaching machine situation plus two conventional situations: making written notes and making a mental outline. The situations evoking the lowest levels of achievement were situations in which the students were reading without writing, planning a talk, and making a written outline.

This succession of situations is not singularly one of difference in materials nor of difference in levels of activity, but is arrayed with respect to the degree to which the activity is structured from moment to moment to moment during study. The student is paced by his progress through the material at hand, regardless of situation, but the freedom he has in dealing with the material varies from little restriction in the low achievement situations to considerable restriction in the high achievement situations. (The results in Situation X remind us that the activity must be relevant, as well as orderly, if the learning is to be effective).

Mean differences in achievement, then, were attributed to the degree to which the responses were structured, rather than to the amount of activity generated. When our quiz questions are studied, it is apparent that high scores will result from convergent rather than divergent thinking, from remembering terminology and relationships;

from seeking out certain logical implications, rather than seeking personal applications or alternate perspectives; from convergent rather than divergent thinking. When the educational objectives measured by our quizzes are pursued (which is most of the time in our schools), it is logical, and consistent with the results of our analyses, that the study situation should be highly structured. Programed materials were used to greater advantage by our students, it was concluded, because of the contribution those materials made to a structured study period.

Several secondary hypotheses were included in the examination of the results because of the availability of data for analysis. Each of the secondary hypotheses are indicated in the following paragraphs with a discussion of the findings.

Retention

Of interest to the investigators was the relationship of retention of lessons learned, with the various conditions of study. The hypothesis was that the different study conditions led to equal retention of acquired learning. The quizzes for the five topics studied first were readministered during the last session. The reader will recall that the first quiz was administered two weeks after the topic was studied. The second administration of the same quiz occurred from five to ten weeks after the first administration. Mean scores are shown in Table 11. Two different analyses were made of the test-retest differences. The results of the first analyses are presented in Table 12.

As expected, there was a significant decrease in the mean quiz scores from the first administration to the second. The insignificant F values for the Retest by Mode, Retest by Condition, and Retest by Mode by Condition supported the null hypothesis that the different study conditions led to equal retention of acquired learning. There were significant differences among the means for Condition of study and the interaction of Mode and Condition with the combined test scores, but the failure to find an interaction of Mode and Condition with Retest indicates that the significant Condition, and Mode by Condition effects obtain in both first and second administrations.

The second analysis of the test-retest effect on quiz scores was an analysis of variance of the loss scores within each of the topics. The results are presented in Table 13. In none of the analyses of loss scores was the Mode of study significantly related to the drop in score from first administration to second. With two of the topics there was a significant difference among the mean loss scores for Condition. The difference for the Insurance Topic is the result of a relatively larger loss from first administration to second for the Programed Text and Teaching Machine Conditions. However these two conditions still had higher means on the final administration than the other three conditions.

The Programed Text and Outline Conditions had the greatest loss

Table 11

Test and Retest Means for 100 Students Studying the
First Five Topics

Condition	Mode	First Test	Retest
Notes			
I	Read; make mental notes	48.88	47.58
II	Read; make written notes	49.16	46.42
Outline			
III	Read; make mental outline	51.80	50.18
IV	Read; make written outline	46.22	42.62
Prepare a Talk			
V	Read; make mental notes	50.50	46.70
VI	Read; make written notes	46.12	44.86
Programmed Text			
VII	Read	52.14	47.68
VIII	Read; write responses	52.70	49.16
Teaching Machine			
IX	Read	52.60	50.84
X	Read, write responses	49.30	47.48
Reading Only		51.19	48.60
Reading and Writing		48.71	46.10
Total		50.00	47.35

Table 12

Analysis of Variance of Retention of Five Topics

Source	DF	Type	Components	SS	MS	F
Between Persons						
Mode	1	F	M + p + E	1423.24	1423.24	2.93
persons in Mode	98	R	p + E	47603.65	485.75	
Within Persons						
Retest	1	R	R + Rp + E	1809.02	1809.02	57.72**
Condition	4	F	C + Cp + E	1984.41	496.10	6.98*
RC	4		CR + CRp + E	288.35	72.09	1.64
Error						
RP						
RM	1		RM + Rp + E	2.51	2.51	< 1
Rp	98		Rp + E	3071.17	31.34	
CP						
CM	4		CM + Cp + E	1898.25	474.56	6.49*
Cp	392		Cp + E	28672.14	73.14	
KCP						
RCM	4		RCM + RCp + E	93.19	23.30	< 1
RCp	392		RCp + E	17209.26	43.90	
Total	999					

** P < .01

* .01 < P < .05

Table 13

Analysis of Variance of Loss Scores by Topics

Source	DF	Sum of Squares	MS	F
(Insurance)				
Mode	1	68.89	68.89	1.54
Condition	4	519.80	129.95	2.90*
Mode X Condition	4	135.16	33.79	< 1
Within	90	4032.90	44.81	
Total	99	4756.75		
(Dewey Decimal)				
Mode	1	57.76	57.76	1.04
Condition	4	488.40	122.10	2.20
Mode X Condition	4	467.24	116.81	2.10
Within	90	4997.60	55.53	
Total	99	6011.00		
(English Money)				
Mode	1	25.00	25.00	< 1
Condition	4	132.70	33.20	< 1
Mode X Condition	4	121.10	30.30	< 1
Within	90	3779.20	41.99	
Total	99	4058.00		
(Logic)				
Mode	1	56.25	56.25	< 1
Condition	4	672.96	168.24	1.72
Mode X Condition	4	413.15	103.29	1.06
Within	90	8798.15	97.76	
Total	99	9940.51		
(Poetry)				
Mode	1	10.89	10.89	< 1
Condition	4	1044.36	261.09	6.72**
Mode X Condition	4	1373.96	343.49	8.84**
Within	90	3496.50	38.85	
Total	99	5925.71		

from first to second administration in the study of poetry. There the means of these two conditions were considerably lower than the means of the other three conditions on the second administration. This result accounts for the significant interaction effect obtained in the poetry analysis. The significant differences among conditions in Poetry and Insurance do not appear to be related to the group involved. That is, different groups were involved in the conditions showing the large loss.

In general the two analyses provide us with no basis for rejecting the hypothesis that different study conditions lead to equal retention of acquired learning. For two of the topics there was a significant difference in loss scores among the conditions. However, our failure to obtain a significant interaction effect between Retest and the other main effects in the first of the two analyses keeps us from rejecting the null hypothesis.

Personal Values and Reaction to Programs

Suggestions have been made in the research literature in education that performance on programed materials might be affected by the personality of the student. A student of a certain personality type might react favorably to the highly structured program, whereas other personality types might react negatively. A crude test of the null hypothesis of no relationship between the value pattern of a student and the condition or mode of study in which his achievement was best was made by correlating the scores of the subjects on each of the areas of the Poe Inventory of Values (PIV) with the achievement scores in each treatment. Had the correlation between value scores and achievement been consistently higher (or lower) for programed instruction conditions than conventional instruction conditions, the null hypothesis would have been rejected. No clear difference can be seen in any of the columns in Table 14.

To look further for a relationship between strong personal values and achievement within different study modes, we found the twenty students with the highest PIV Scores in each value area. We looked at their achievement scores in each of the ten instructional situations. We compared their achievement to the achievement of all students. Each row of Table 15 indicates how different a group of 20 high value students was from the entire group of 100 students. The 23 in the upper left cell in Table 15, when scaled back to 2.3, indicates that Situation I mean achievement for the 20 students who were highest on the PIV Intellectual Values Scale was 2.3 (T scale) points higher than the corresponding mean achievement for all 100 students.

The consistency of deviation scores across the rows of Table 15 merely means that 3 scales, Intellectual, Power, and Material, yielded scores positively correlated with achievement whereas 1 or more other scales, notably Humanitarian, yielded scores negatively correlated with

Table 14
The Correlation between Achievement and Personal Values as
Found in Ten Instructional Situations

		Value							
		Aes- thetic	Intel- lectual	Material	Power	Social Contact	Religious	Pres- tige	Humanit.
	I	1	2	0	1	-2	-3	-1	-3
T	II	0	2	1	1	-1	-3	0	-1
R	III	0	0	0	2	-1	-2	-1	-2
E	IV	-1	2	-1	0	-2	-2	-3	-1
A	V	0	2	1	1	-1	-2	-1	-1
T	VI	1	2	-1	1	-2	-3	-1	-3
M									
E	VII	0	2	0	0	-2	-3	-1	-2
N	VIII	2	2	0	0	-1	-3	-1	-1
T	IX	-1	2	1	1	-1	-3	-1	-2
X	X	0	1	0	0	-1	-1	-2	-2

Required for significance: 20 at 5% level and 26 at 1% level.

Decimal points omitted.

Table 15

Mean Deviations* in Achievement of Students with High Value Scores
from Achievement Means for All Students

Students with High Value Scores on Scale	Situation										All Ten
	I	II	III	IV	V	VI	VII	VIII	IX	X	
Intellectual	23	41	30	51	13	49	37	27	27	23	32
Power	13	24	42	10	36	40	18	35	39	21	28
Material	11	38	11	19	27	11	19	16	11	16	18
Aesthetic	16	-4	11	-33	-14	28	-26	32	25	-4	-2
Prestige	-30	-12	-36	-8	5	-12	-6	4	-12	-10	-11
Social	0	-34	-7	-44	18	-14	-21	-10	-1	-8	-12
Religious	-16	-24	-14	-4	10	-26	-19	-14	-19	-14	-14
Human'n	-32	-21	-26	-48	-29	-35	-41	-31	-42	-68	-37

* Each mean deviation scaled by 10 to eliminate decimal point. Means for all 100 students shown in Table 5.

achievement. If there had been differences in deviations to the left and right of the vertical line separating programmed situations from nonprogramed situations we would have rejected our null hypothesis. In this data we could see no evidence of an interaction of value orientation with programing of instruction. Whether or not a person achieved more or less with programing did not appear to be related to his responses on a personal values inventory.

An indication that values might not be related to propensity toward programmed instruction was obtained from the data in Table 16. In the last session the students responded to the question, "Do you find anything objectionable about studying with teaching machines?" The yes-no responses to this question were tabulated against score ranges of ten for each area of the PIV. There is some indication that persons who are high in intellectual and religious values tend to object more to teaching machines, but the relationship does not appear to be strong.

Table 16

Numbers of Students with Various Value Scores Objecting
to the Use of Teaching Machines

	PIV Score													
	0-19		20-29		30-39		40-49		50-59		60-69		70 & up	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Aesthetic	7	6	16	14	17	12	12	6	5	5	0	0	0	0
Intellectual	0	0	3	3	9	7	20	19	14	11	10	3	1	0
Material	14	12	19	13	19	14	3	3	2	1	0	0	0	0
Power	3	3	2	6	15	5	18	15	7	10	11	3	1	1
Soc. Contact	1	2	4	2	6	1	20	12	17	14	8	10	1	1
Religious	4	5	8	7	21	16	17	14	7	1	0	0	0	0
Prestige	6	4	8	9	23	15	13	11	6	4	1	0	0	0
Humanitarian	0	1	2	1	3	6	23	16	23	11	4	7	2	1

Opinions on Instructional Materials

The students were asked which topics they liked most and least well and which topics they learned most and least about. The tabulations for these items are in Table 17. Although the topics liked best were also those learned best, as indicated by the group, there was no apparent tendency for an individual to indicate that the topic he felt

Table 17
Numbers of Students Preferring Each of Ten Topics

	Liked Most	Liked Least	Learned Most	Learned Least
Insurance	8	12	13	7
Dewey Decimal	2	12	7	2
English Money	7	10	12	8
Poetry	19	6	11	8
Logic	3	8	3	4
Nebr. Government	11	12	15	9
Leaves	5	22	2	30
German Money	10	3	13	7
Art	27	3	16	9
Fermentation	8	12	8	16

he learned most was also the topic he liked best. There are eleven persons registered in the major diagonal of Table 18. The other 89 students each indicated a learned-most topic different from his liked-best topic.

Table 18

The Topic Liked Most and the Topic Learned Best as Indicated by
100 Students

	Topic Learned Best									
	Ins	Dec	E M	Poet	Log	Govt	Leaf	G M	Art	Ferm
Insurance	2	2	0	1	0	0	3	0	0	0
Dewey Dec	0	0	0	1	0	0	0	0	1	0
Eng Money	0	1	0	2	2	0	1	0	0	1
Poetry	2	2	1	2	5	0	1	0	0	6
Logic	0	1	0	1	0	1	0	0	0	0
Nebr Govt	1	3	3	1	0	0	2	0	1	0
Leaves	0	0	1	0	0	0	1	2	0	1
Ger Money	1	0	0	1	2	0	1	0	5	0
Art	2	3	2	3	2	0	1	0		0
Ferment 'n	0	1	0	0	0	2	2	0	2	1

Data were tabulated in Table 19 to determine whether feelings about the topic were related to the study condition. Relationships were significant with three of the four classifications at the .10 level. The topic the student liked least tended to be one studied in a situation less conducive to high achievement. The topic the student believed to be the one he learned most about tended to be one studied in a situation more conducive to high achievement. The topic the student judged least well learned tended to be one studied in a situation actually less conducive to high achievement.

Another indication of the relationship between topic liked best as judged by the subject and topic learned best as revealed by the criterion score was obtained by ranking the achievement scores for each subject. The subject's highest achievement score was given the rank of one, second highest two, and so on. The mean achievement test rank for the topics believed liked best by the subjects was four. A mean rank of four was indicative of a slight tendency for the subjects to perform better on the topics liked best.

Table 19

Student Preferences for Topics
Classified as to Conduciveness to Achievement
For the Whole Group

	<u>Topics Studied Under Conditions</u>			Chi Square
	Most Conducive To Achievement (VII,VIII,IX)	More Conducive To Achievement (X,III,II)	Least Conducive To Achievement (I,IV,V,VI)	
Students Selecting the Topic Liked Best				
Actual frequency	36	30	34	2.10
Expected frequency	30	30	40	
Students Selecting the Topic Liked Least				
Actual frequency	16	33	51	9.85*
Expected frequency	30	30	40	
Students Selecting the Topic Learned Best				
Actual frequency	38	35	27	7.18*
Expected frequency	30	30	40	
Students Selecting the Topic Learned Least				
Actual frequency	20	32	48	5.06*
Expected frequency	30	30	40	

*Significant at the .10 level

When topics liked most and least or the topic believed learned best or least were compared with mode and condition of study the only relationship significant at the .05 level was between topic liked least by study materials. There was a tendency for the topic liked least to be associated with conventional materials. As shown in Table 20, three other contingencies were found significant at the .15 level.

Table 20

Number of Students Preferring Topics Classified as to Response Type and Study Material Type

	A Topic Which the Student Read and Thought About	A Topic Which the Student Read and Wrote About	Chi Sqr	A Topic Which Was Read in Conven- tional Format	A Topic Which Was Read in Programed Format	Chi Sqr
STUDENTS SELECTING THE TOPIC LIKED BEST						
Actual frequency	55	45	1.00	52	48	2.67*
Expected frequency	50	50		60	40	
STUDENTS SELECTING THE TOPIC LIKED LEAST						
Actual frequency	48	52	.16	70	30	4.17*
Expected frequency	50	50		60	40	
STUDENTS SELECTING THE TOPIC LEARNED BEST						
Actual frequency	42	58	2.56*	52	48	2.67*
Expected frequency	50	50		60	40	
STUDENTS SELECTING THE TOPIC LEARNED LEAST						
Actual frequency	49	51	.02	66	34	1.50
Expected frequency	50	50		60	40	

*Significant at the .15 level

Attitude Toward Machines

Additional information was obtained indicating the students' reaction to programed materials. Some negative feelings about programed materials and teaching machines were evident from the data of Tables 21 and 22.

Fewer subjects at the end of the experiment felt it to be desirable to replace most or all conventional texts than at the beginning of the experiment. If no significant change from the beginning to the end of the experiment is hypothesized, the data in the upper half of Table 21 yield a chi square of 3.25. With one degree of freedom, such a chi square value's probability is $.05 < p < .10$.

Table 21

Changes in Student Reactions to Programed Materials
During the Study

How much (sic) of your regular textbook materials do you think would be more effectively studied if they were programed materials?

	Number of Students Responding	
	"None" or "A Few"	"Most" or "All"
Pretest	55	45
Posttest	64	36

Do you find anything objectionable about studying with teaching machines?

	Number of Students Responding	
	"Yes"	"No"
Pretest	38	62
Posttest	57	43

The chi square value for the bottom half of Table 21 is 15.62 when no change is hypothesized from pretest to posttest response to the question. The probability associated with this chi square is $p < .001$. Many more subjects found something objectionable about teaching machines at the end of the experiment than at the beginning. On the other hand, approximately 80 per cent of the students on both administrations of the questionnaire responded that programs should be available as supplementary material, that teaching machines should be available, and that they would use teaching machines in the classroom if they were teaching. It appeared the students did perceive some value for programs and teaching machines as supplementary material, but that their experience with programs and teaching machines was such that they were not enthusiastic about the prospects of completely automated teaching.

The investigators believed much of the negative reaction to the teaching machine was caused by the mechanical problems that came up quite regularly while the students were using them.

Table 22

Group Means and "F" Values of 5-Item Pretest and Posttest of Each Topic

Pretest	Group										
	A	B	C	D	E	F	G	H	I	J	"F"
Insurance	0.7	1.5	1.0	1.3	0.8	1.1	1.6	1.3	1.0	0.6	1.12
Dewey Decimal	1.1	1.5	1.5	1.2	1.4	1.9	1.3	1.8	2.8	2.3	2.34*
English Money	0.8	0.8	0.8	0.8	0.7	0.8	0.9	0.9	1.1	0.7	1
Poetry	1.3	1.1	0.9	0.8	0.7	1.0	1.3	1.1	1.4	1.8	1
Logic	0.1	0.3	0.3	0.0	0.5	0.2	0.1	0.3	0.2	0.3	1
Government	0.9	1.8	1.2	1.3	0.9	1.8	1.9	1.8	2.1	1.6	2.35*
Leaves	0.5	0.8	0.7	0.8	0.2	0.7	1.1	1.1	1.6	0.5	1.30
German Money	0.3	0.5	0.6	0.8	0.7	0.7	0.3	0.4	0.4	0.3	1.00
Art	1.6	1.6	1.9	1.4	1.4	1.2	1.7	1.8	1.6	1.9	1
Fermentation	0.2	0.4	0.1	0.4	0.4	0.7	0.5	0.6	0.2	0.1	1
Grand Means	0.8	1.0	0.9	0.9	0.8	1.0	1.1	1.1	1.2	1.0	
Verbal	45.0	56.8	55.6	44.8	54.7	56.5	47.9	32.1	58.1	63.1	1.29

Note.--With 9,90 degrees of freedom, $F_{.05} = 1.95$, $F_{.01} = 2.62$.

Summarizing the results of the tests of the secondary hypotheses: No relationship was found between the activity (mode or condition) of study and retention; No relationship was evident between value patterns and study conditions or feelings about using teaching machines. Some relationship was evident between topics enjoyed most or least and topics believed learned best or least well. Some negative change was observed in regard to feelings about use of programmed materials as complete course materials and the use of teaching machines, but a majority of the students did express a belief that machines and programmed materials can be useful.

Initial Equivalence of Groups

Data on the equivalence of groups are presented in Table 22. With respect to pretest achievement items, there was a significant difference among group means on the Nebraska Government and Dewey Decimal System sections. For the other eight topics the "F" value was less than unity or only slightly more.

The group means for Nebraska Government were proportional to the grand pretest means, so the significance there can be attributed to differences in general knowledge among the groups. Achievement on the Dewey Decimal topic was not highly correlated with pretest achievement on the other nine topics. Therefore, it was concluded that there was some disparity of initial acquaintance with this topic.

The groups with advantage here were E, F, I, and J, which had been assigned Situations III, IV, VII, and VIII for the study of the Dewey Decimal System. Achievement on this topic measured on the two week quizzes did run much above expectation for Group J and slightly above expectation for Groups E and I, but much below expectation for Group F. Group J's situation, Situation VII, may have been credited with higher achievement which was actually due to specific knowledge of students in that group. But this was but one instance out of one hundred where we might have found initial disparity between groups affecting the level of achievement. We concluded that the groups were adequately matched.

Differences among the groups as to verbal ability were not found to be significant.

CHAPTER IV

AN ANALYSIS OF IMPACT

The quite general problem to be considered in this chapter is the problem of scaling a set of experimental conditions or environmental influences. We wish to scale these conditions or influences in order to describe them more precisely, and, hopefully, in order to understand them better. We are concerned about the impact of each of several conditions on a single criterion variable. To emphasize this fact, we will call the scaling procedure to be developed here, impact analysis.

In the usual scaling problem all subjects are exposed to a set of stimuli. Each subject perceptually compares and implicitly quantifies the stimuli. The data are analyzed to yield scale values which reflect the subjective value that this group of subjects associates with each stimulus.

With impact analysis all subjects are exposed to a set of stimulus conditions. Each subject responds while or after being exposed to each condition, and the level of criterion performance attained is recorded. The data are analyzed to yield scale values which reflect the impact potential of each condition, the potential for altering individual performances.

For a given set of stimulus conditions, one or more sets of scale values will be obtained. When multiple scale values are assigned to each condition, the solution is said to be multidimensional. Each dimension is a different basis for describing the potential impact of stimulus conditions. For each additional dimension considered, the disparity between observed and "predicted" levels of performance will be at least somewhat reduced. A study of the scale values and the reduction in error which accompanies each additional dimension should enlighten us as to the uniquenesses and complexities among the experimental conditions.

An Emphasis on Individual Differences

This impact scaling method is based on the premise that persons react differently but systematically to impinging conditions. Reaction is considered here as a personal phenomenon rather than a group phenomenon. The basic functional relationship (given in Equation 4) accounts for variation of scores within individuals as well as variation across individuals. Reaction or level of performance will be expressed for each individual as a function of his own personal characteristics interacting with external influences.

During experimentation, level of performance is the dependent variable. For impact analysis level of performance is the independent variable in the sense that it is input. Scale values for persons and for conditions are dependent variables, in the sense that they are the ultimate output. Both personal and situational characteristics are scaled, but the primary objective is one of obtaining the scale values associated with each experimental condition.

To delimit the circumstances in which impact analysis might be used it is useful to distinguish between nested and crossed experimental conditions. We are considering only those conditions that alter individual output. Conditions nested in persons; e.g., sex, personal commitment, physical configuration, aptitude; cannot have an impact in this sense on personal performance, and therefore are outside the domain of impact analysis. Crossed conditions; e.g., stimulus pattern, rate of presentation, climate, reward, and mode of operation; are conditions in which all subjects can participate or to which all can be exposed. Only crossed conditions can be studied with impact analysis.

During the scaling itself we are not attempting to compare groups of individuals. After important dimensions or clusters of conditions have been identified, however, we are likely to want to compare groups as well as to describe the individuals and environments involved.

Functional Scales

Using rational and traditional frames of reference we often ignore the fact that two conditions may be logically independent, but yet be functionally interrelated. Suppose that for every student whose achievement drops a few units because the teacher is hypercritical, this student's achievement is enhanced half that many units when lighting is improved. Teacher personality and lighting are presumed to be experimentally independent conditions, but it is possible that the susceptibility of students to them is proportional. For some purposes it would be useful to treat any two or more functionally interdependent conditions, whether or not they are logically compatible, under the same rubric. Proximal scale values from impact analysis indicate similarity of impact rather than logical affiliation.

A parsimonious description is attained when all pertinent conditions are scaled unidimensionally and represented as points on a single continuum. Such representation would be sufficient if it were found that all conditions were related in such a way that the impact of each condition is a simple function of the impact of any other. Such simplicity is unlikely in our world, however. If several continua are needed to properly represent the different conditions, the set of conditions is multidimensional.

As a result of the analysis of impact each dimension is represented by a separate scale. Every condition has a scale value on each

dimension. If a particular scale value, X_{kj} , is found to be about zero, then the impact of condition j is negligible as far as dimension k is concerned.

A Linear Regression Model

The model for this analysis is an expression in terms of multiple regression. Hyperplanes are generated separately for each individual, but are interrelated through a common set of impact scale values, X . Performance, Y , is seen as a personal response; varying differently from person to person, but systematically for a given person, as conditions vary.

It is obvious that on any given occasion an individual performs at a level which differs at least somewhat from that which is "characteristic" or "expected" of him. The major premise of this model is that the amount of difference is a function of personal susceptibilities to nonstandard conditions and of degrees to which the conditions influence everybody's performance. In other words, the change in level of individual performance for a given condition is attributable to interaction between the person's susceptibility and the relative impact of the particular condition, both calculated for each dimension. With three dimensions of impact, an individual's performance in one situation would be given by

$$Y_{ij} = Y_i' + a_{i1}X_{1j} + a_{i2}X_{2j} + a_{i3}X_{3j}, \quad (1)$$

where

Y_{ij} is the level of performance by the i -th person when exposed to the j -th condition,

Y_i' is the characteristic level of performance for the i -th person,

a_{i1}, a_{i2}, a_{i3} are the coefficients of susceptibility for the i -th person with reference to the first three dimensions of impact,

X_{1j}, X_{2j}, X_{3j} are the impact scale values for the j -th condition with reference to the first three dimensions of impact,

$i = 1, 2, 3, \dots, N$ persons, and

$j = 1, 2, 3, \dots, n$ conditions.

The level of performance which we consider characteristic for an individual might vary from one broad category of conditions to another.

The characteristic performance designated by Y_i^j should refer to the category of conditions represented by the conditions under investigation. A reasonable estimate of this characteristic level of performance is $\bar{Y}_i$, the mean performance for individual i , averaged across all observed conditions. Substituting $\bar{Y}_i$ for Y_i in (1) we have

$$Y_{ij} = \bar{Y}_i + a_{i1}X_{1j} + a_{i2}X_{2j} + a_{i3}X_{3j} \quad (2)$$

Subtracting this mean from both sides of the equation,

$$Y_{ij} - \bar{Y}_i = a_{i1}X_{1j} + a_{i2}X_{2j} + a_{i3}X_{3j} \quad (3)$$

Equation 3 indicates that impact, measured as a change in level of individual performance, is a function of susceptibility and impact potential.

The impact, that change in performance specified in (3), is attributed to the condition as a whole rather than to certain of its parts. Three conditions that differ in only one characteristic by proportionate differences will not necessarily have scale values so proportioned. The question of what critical parts or features of a condition cause it to have such an impact is left for later consideration.

A more general expression of the model, i.e., for k dimensions is

$$Y_{ij} = \bar{Y}_i + \sum_{k=1}^m a_{ik}X_{kj} \quad (4)$$

It should be noted here that $m \leq n$ because the variance of Y can always be completely accounted for in $n-1$ dimensions.

As expressed in (4) the relationship between performance on the one hand, and personal susceptibility and impact of conditions on the other, is represented as a linear relationship. A nonlinear model may later be preferred, either for rational or empirical reasons, but for the time being this simple function will be used.

Expressed in matrix notation Equations 3 and 4 become

$$(Y - \bar{Y}) = AX \quad (5)$$

Errors of Fit

Equations 1-5 represent an idealized relationship between performance, impact potential, and personal susceptibility. With real data the Y scores generated from a solution of A and X will

almost certainly differ from the observed Y scores. The observed scores should be related to scale values by

$$(Y-\bar{Y}) = AX + E, \quad (6)$$

where E is a matrix of errors of fit. An element of E is the discrepancy between an observed score and the corresponding score given by the regression equation; i.e.,

$$e_{ij} = Y_{ij} - \bar{Y}_{i.} - \sum_{k=1}^m a_{ik} X_{kj} \quad (7)$$

The applicability of this model for any number of dimensions will be indicated by the size of the sum of squares of errors,

$\sum_{i=1}^N \sum_{j=1}^n e_{ij}^2$. A solution is obtained when this grand sum is a minimum for all possible combinations of a_{ik} and X_{kj} .

Solving for Parameters

The problem of estimating personal characteristics a_{ik} would be a straight-forward regression problem if the X_{kj} values were known. On the other hand, if the personal characteristics were known, the condition scale values, X_{kj} , could be solved by matrix algebra. In circumstances where impact analysis is to be used, however, neither X_{kj} nor a_{ik} values are known. A solution is possible by factoring the square symmetric $(Y-\bar{Y})(Y-\bar{Y})'$ matrix.

According to the theorems of Eckart and Young (1936) there is a matrix F of the same order but of smaller rank than the matrix of observations $(Y-\bar{Y})$. The product of this F matrix and its transpose, $F'F$, approximates the product $(Y-\bar{Y})'(Y-\bar{Y})$ in a least squares sense. The analysis of the F matrix can be considered an analysis of the $(Y-\bar{Y})$ matrix.

According to the procedures of Hotelling (1933) and others (Horst, 1963; Tucker and Messick, 1963), an F matrix can be analyzed as follows

$$F' = UDW \quad (8)$$

where

U is an $n \times r$ orthogonal matrix ($U'U = I$),

D is an $r \times r$ diagonal matrix of latent roots, and

W is an $r \times N$ orthogonal matrix ($WW' = I$).

The U matrix is a solution for our A matrix and the W is a solution for our X matrix.

As in many similar procedures such as factor analysis a solution for A and X which provides a least square fit to the $(Y-\bar{Y})$ matrix is not a unique solution. Many combinations of A and X will give a least squares fit. The principal components solution is one of these solutions that is unique in that it maximizes the variance accounted for by successive dimensions. It is the principal components solution that will be obtained in our analysis of impact. Some rotation of the scale values, however, may provide an arrangement of dimensions more in keeping with our ideas about the experimental conditions.

Solving for Dimensions

A major question is that of deciding on the number of dimensions that should be interpreted. To use only a single dimension to describe influences is parsimonious, but the impact of the different conditions may be inadequately described. On the other hand, the immediate data can be completely "explained" by referring to as many dimensions, less one, than the number of conditions. But in so doing, of course, simplicity is abandoned. The impact indicated by most of the dimensions is probably best attributed to chance.

We wish to describe as much of the variability in performance as can be attributed to dimensions that are reproducible. Reproducible variance is often indicated by the statistic r^2 .

$$r^2 = 1 - \frac{\sum_{ij} e_{ij}^2}{\sum_{ij} (Y_{ij} - \bar{Y}_{..})^2} \quad (9)$$

The r^2 attained by considering k dimensions can be compared to the r^2 for $k-1$ dimension. When each additional dimension appears to be increasing the reproducible variance by the same proportion we will presume that the additional dimensions do not warrant interpretation.

Graphic Examples of Impact

Suppose that we have four learners studying under two conditions. Their achievement is represented by the marks on the first two vertical lines in Figure 9. The points for each person belong to a family of points, i.e., a hyperplane, which represents the relationship between the study conditions and achievement for that person. We are assuming that this relationship is linear. We pass a straight line through the intercepts for Conditions I and II. Now let the learners study in a

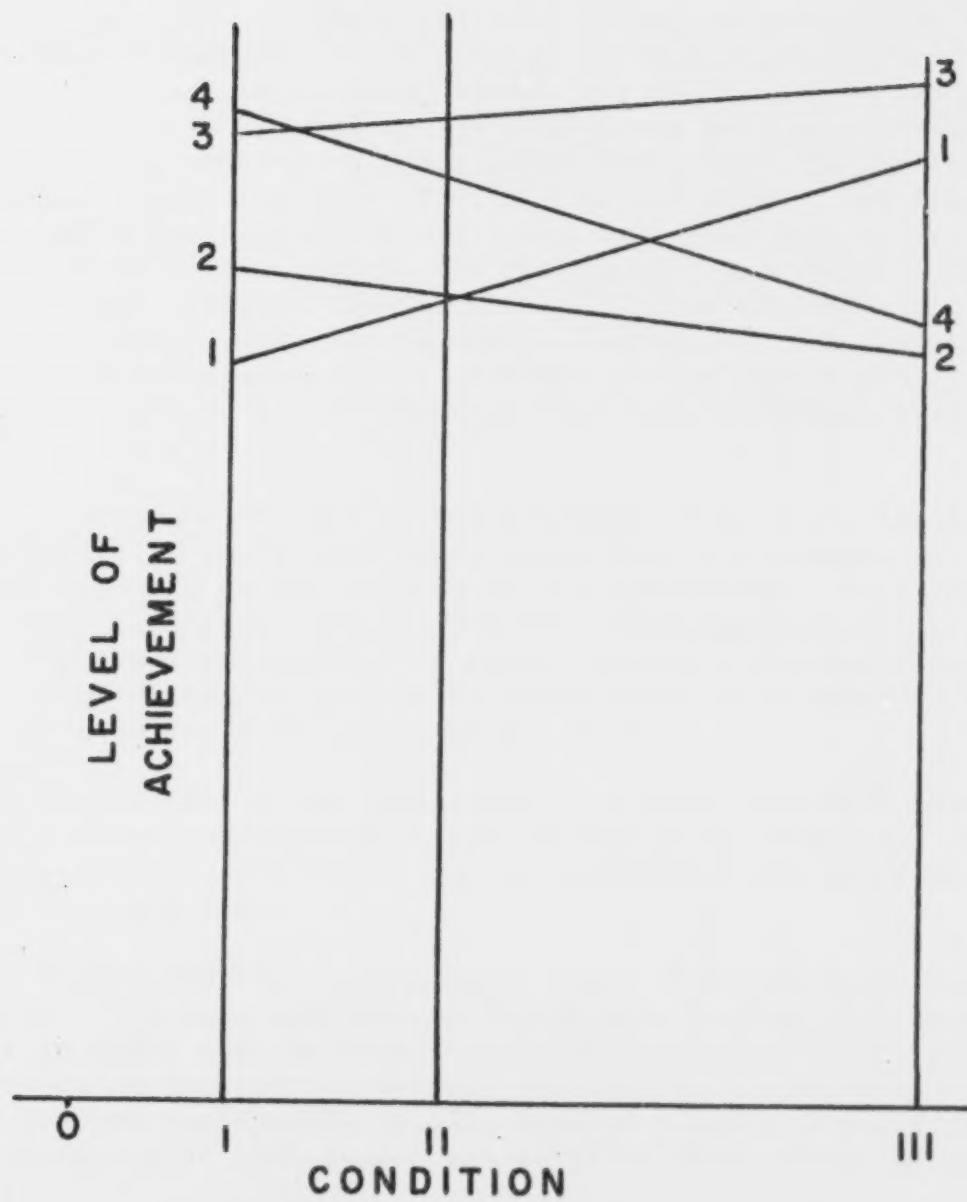


Figure 9. Unidimensional representation of achievement by 4 learners in 3 conditions.

third condition. If our assumption of linearity is sound and if the conditions can be scaled unidimensionally, we can erect the Condition III ordinate somewhere so that it intercepts the performance lines properly. In Figure 9 the third ordinate is so located.

Multidimensionality occurs when several different bases are needed for categorizing the conditions as to the impact they have on individual performances. In a general sense multidimensionality is similar to interaction between persons and conditions in analysis of variance. There is a considerable difference however. Anova interaction of persons with conditions means that there are reliable realignments of persons from condition to condition. For that analysis, how the realignment occurs is not important, just that it recurs. According to the impact model, the realignment must reflect systematic personal change. Impact dimensionality, either unidimensional or multidimensional, means that each person is responding to different conditions as if his response were determined by the impact potential of the condition and by his own susceptibilities, invariant over conditions.

If the conditions are multidimensional then a unidimensional representation of individual achievement does not account for performances observed in one or more of the conditions. Performances in these latter conditions may not fit the unidimensional solution because overall level and variability of performance are incompatible with that in other conditions, or because in these conditions there is a realignment of persons (as with interaction), or both.

In Figure 10a, three conditions have been used with four subjects. The levels of performance vary from condition to condition. The scale is unidimensional. All points for an individual can be connected by a single straight line.

In Figure 10b, four individuals react differently to the three conditions. The mean achievement for persons varies just as much in (10b) as in (10a) and the mean square for conditions is also unchanged but there is an interaction between persons and conditions. Persons E and G are not susceptible at all; Persons F and H are affected. But the performances of each person can still be represented unidimensionally.

In Figure 10c, the sum of squares for interaction, as well as main effects, are the same as in (10b). But the three conditions cannot be arranged so as to permit a straight line representation of each person's performance. A second dimension is required if the model is to fit the data perfectly.

Thus, if there is interaction between persons and conditions the scales may be unidimensional or multidimensional. If there is no such

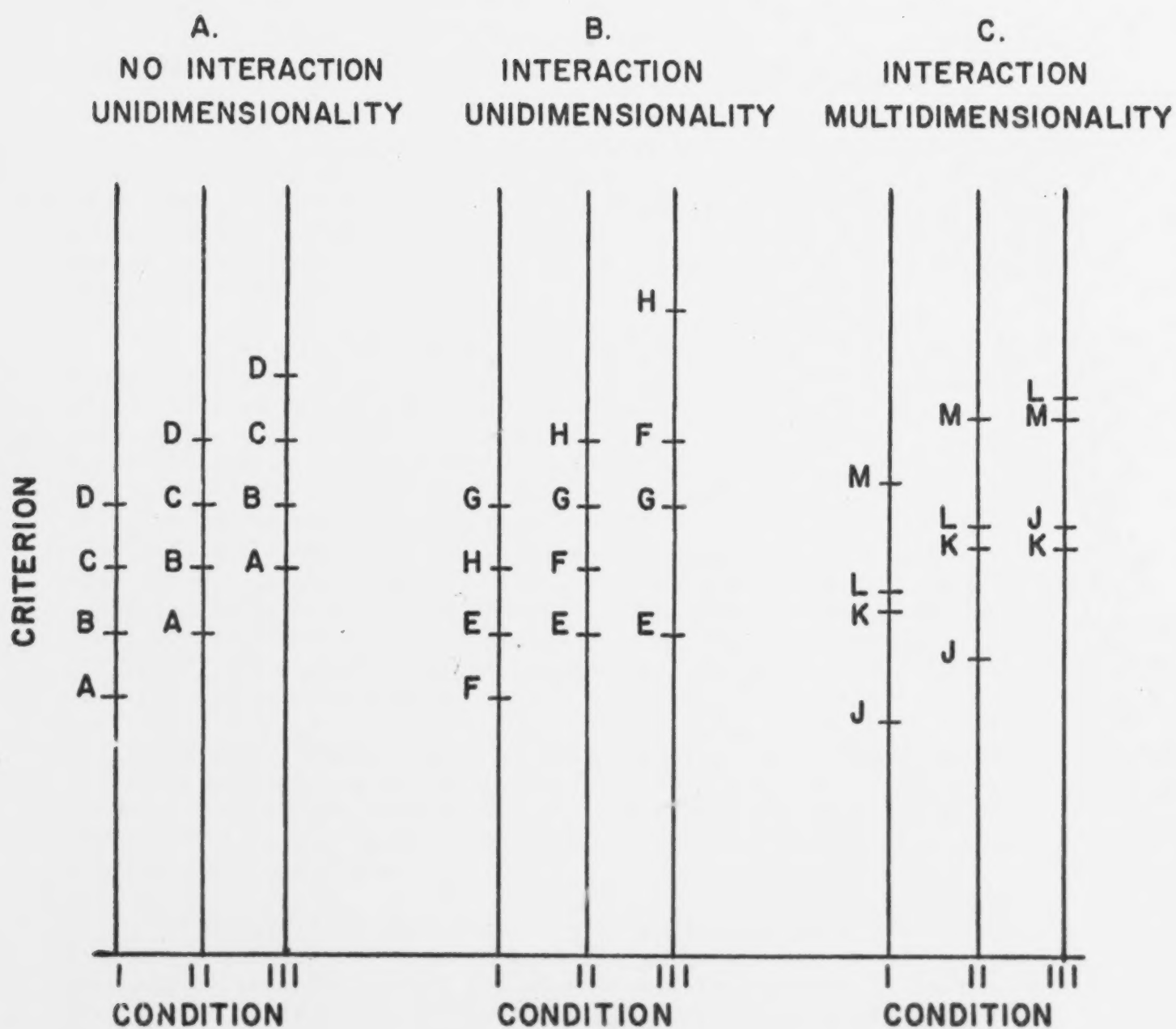


Figure 10. Criterion performances of 4 individuals without interaction, 4 individuals with interaction but without multidimensionality, and 4 individuals with both interaction and multidimensionality.

interaction, the space can have no more than one nonerror dimension. If it is possible to test for this interaction, we perhaps should do so before carrying out an impact analysis. In our data it was not possible to test the interaction between persons and conditions.

The Distinction between Impact and Correlation Analyses

Dimensionality, as defined here, is not based on the correlation of scores. In the unidimensional situation of Figure 9, the correlation is 0.89, -0.06, and 0.41. Elsewhere the coefficients will range from 1.00 to -1.00. Conditions with proximal scale values are going to have highly correlated scores, but those with distal scale values may be highly correlated. Correlation is not sufficient to determine scale placement or dimensionality.

In Figure 11, Condition II is correlated with nothing whereas Condition I and III have a perfect negative correlation. A fourth condition located halfway between (I) and (II) would have a perfect correlation with (I) in this particular set of conditions. A fourth condition can be added to those in Figure 11, correlating perfectly with a condition there without sharing the existing dimension. This is shown in Figure 12, where Condition IV is perfectly correlated with (I), but because of its different mean and standard deviation, requires a second dimension. There is some constant which could be added to (IV) scores to make the situation unidimensional, but the level of criterion scores is presumed to be important and non-arbitrary. There is no justification for ignoring it here as we would in a correlational analysis.

The means are important in this analysis, but neither equal nor different means are a prerequisite for dimensionality. In Figure 11 the conditions means were different. In Figure 13 the means of the four conditions are equal. Yet in both figures there is a unidimensionality of conditions.

The analysis of dimensionality proposed here is different from a factor analysis, both in concept and in purpose. The usual factor analysis is an inquiry into the dimensionality of tests. The resulting factors are scales on which we indicate the similarity with which several tests discriminate among examinees. Impact analysis is an inquiry into the dimensionality of influences on a single criterion performance. The resulting dimensions are scales on which we can indicate the potential of a condition for altering performance.

Impact Analysis of the Achievement Data

The 1000 criterion scores obtained in this study were subjected to impact analysis. We wanted to see how much we could reduce the error variance by employing one or more impact dimensions. From the

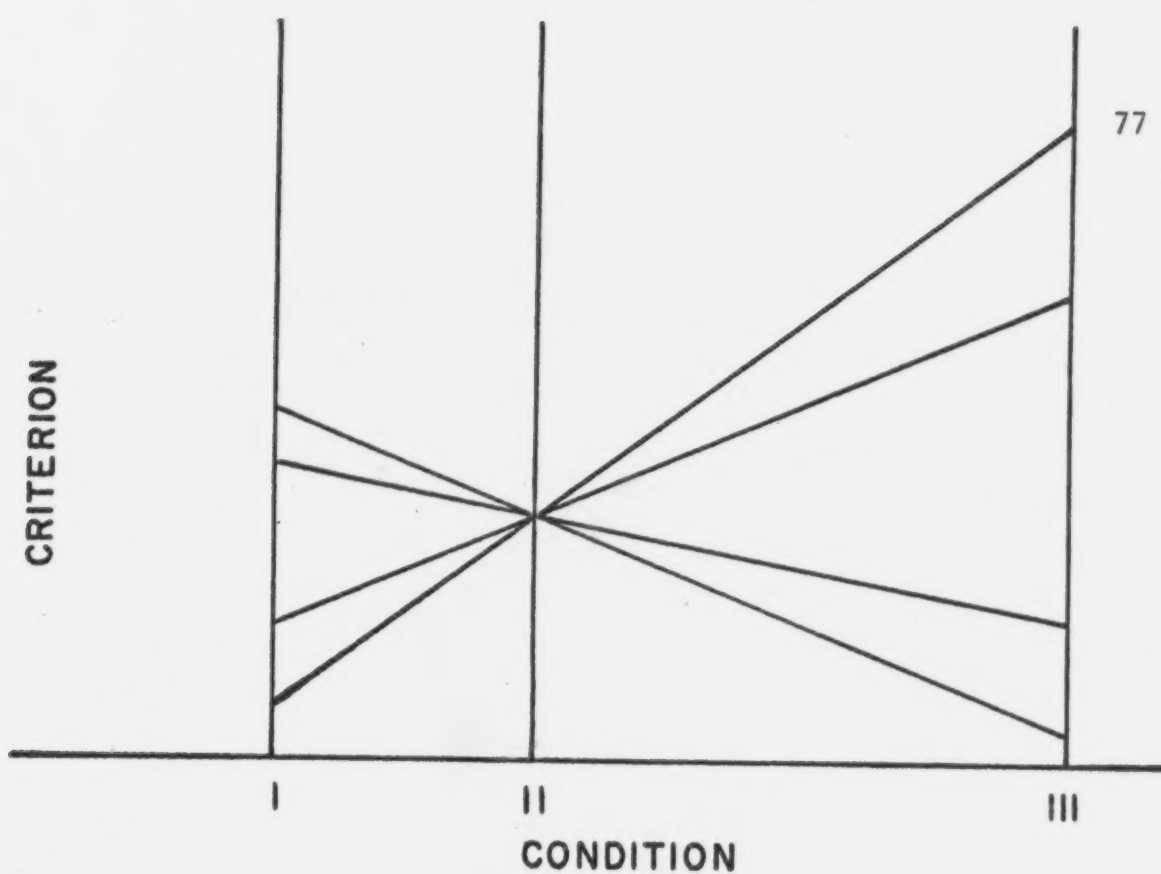


Figure 11. Unidimensional representation of achievement of 4 individuals.

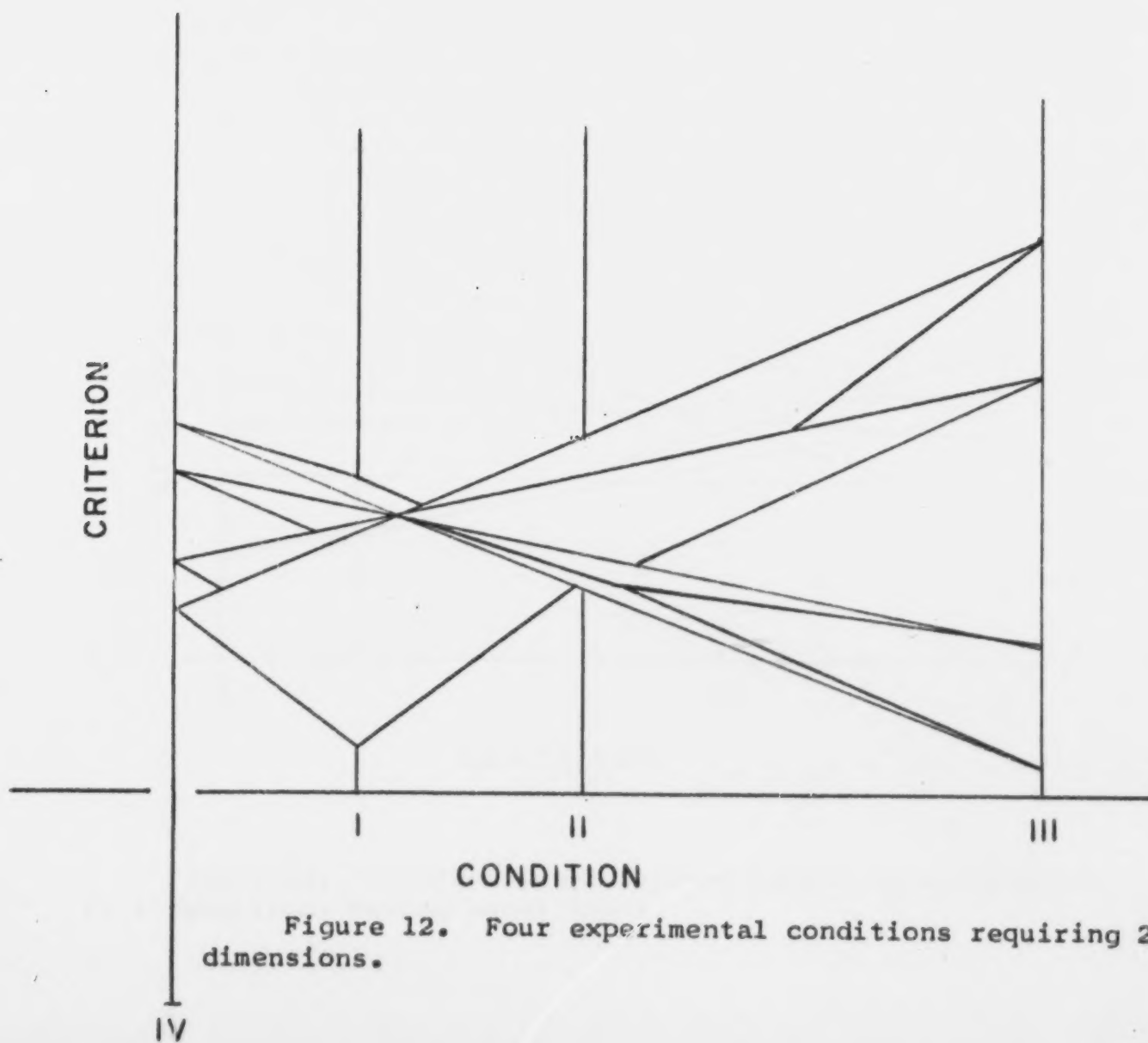


Figure 12. Four experimental conditions requiring 2 dimensions.

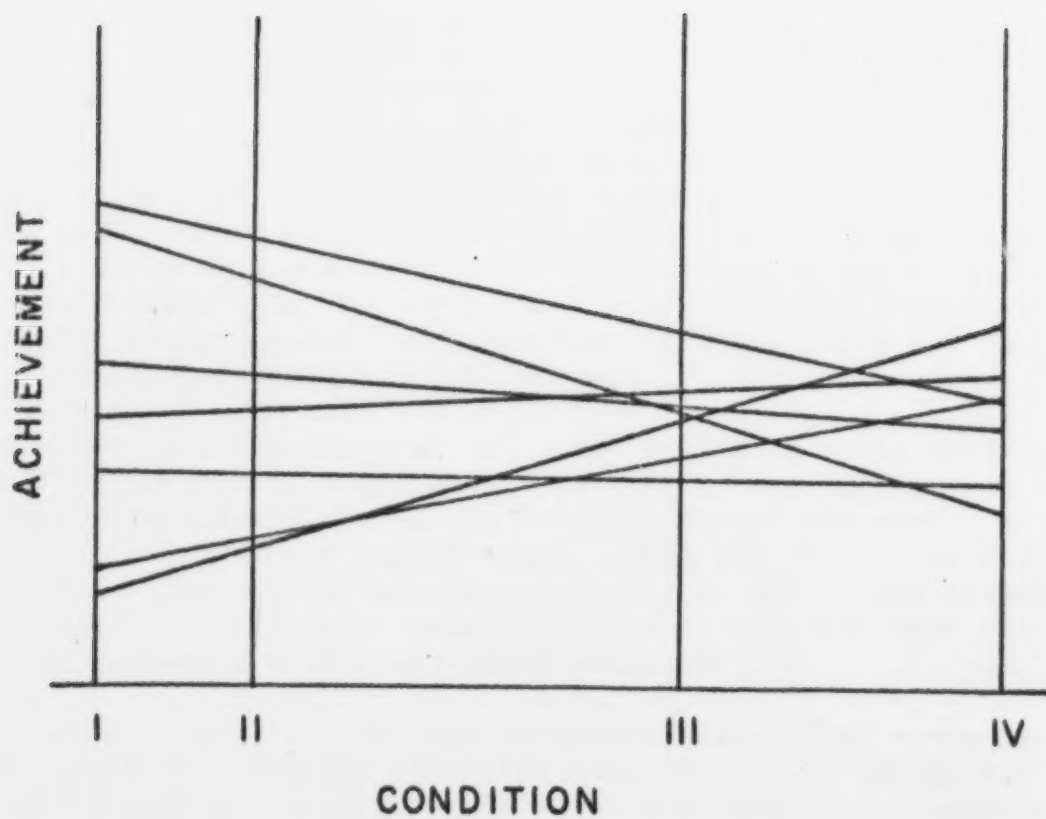


Figure 13. Unidimensional representation of achievement in 4 conditions having equal means.

analysis of variance we acknowledged that the sum of squares of errors was 52040.000 when the main effect of persons was extracted. The same sum of squares would be found if the students were not susceptible to the different conditions (if $a_{jk} = 0$) or, the equivalent, if each condition had zero impact (if $x_{jk} = 0$). In the analysis of variance when both persons and situations were extracted the sum of squares of errors was reduced to 48682.500.

With a single dimension of impact the error sum of squares was reduced to 43150.319. As shown in Table 23 this reduction raised the r^2 to .58 from the .49 obtained when only person means were considered. The increases in r^2 down (in Table 23) to the 9th dimension are quite similar, however. Knowing that at least some of the dimensions are error dimensions persuaded us to treat all ten as error dimensions.

Although the jump from .49 to .58 seems to be a hefty increase it must be recognized that an additional 100 parameters are committed for each dimension utilized. If the commitment for the unidimensional solution were considered equivalent to a loss of 100 degrees of freedom out of 800 remaining, the ratio of mean squares for the gain would be $(52040.000 - 43150.319)(800)/(43150.319)(100) = 1.6$. With 100 and 800 degrees of freedom, an F greater than 1.4 occurs by chance less than five percent of the time. On this basis the increase might be considered statistically significant.

The analysis of variance model, however, provides us with an explanation based on both person and situation means. The sum of squares of errors remaining after this explanation was 48682.500. Reduction of error to 43150.319 increases the r^2 from .52 to .58. Using the reasoning of the previous paragraph a ratio of mean squares of only 1.1 would be associated with the advantage of the unidimensional impact model over the analysis of variance model. The use of impact scale values instead of person and situation means was not supported by this approximate statistical test.

As shown in the columns of r^2 and partial r^2 the variance of the $(Y-\bar{Y})$ matrix is quite evenly distributed among the nine dimensions. The orderliness of the original scores was pretty well removed by subtracting the person means. When the Y matrix was substituted for the $(Y-\bar{Y})$ matrix and factored, 98% of the variance was associated with the first dimension. It is apparent that these data do not require a multidimensional solution.

For what it may be worth the unidimensional impact scale is shown in Figure 14. The ten situation means and scale values are identified by Arabic rather than the Roman numerals used elsewhere. Those first two scales have a rank order correlation of +.7 but, as we have stressed, they provide different information. The impact

Table 23

Reduction of Error Variance in Achievement Scores by Analysis of
Variance and by Impact Analysis

Error	k	$\sum_{ij} \sum e_{ij}^2$	Parameters	r^2	Partial r^2
$e = Y_{1j} - \bar{Y}_{..}$		101674.000	1	0	0
$e = Y_{1j} - \bar{Y}_{.j}$		98316.500	10	.03	.03
$e = Y_{1j} - \bar{Y}_{1.}$		52040.000	100	.49	.49
$e = Y_{1j} - \bar{Y}_{.j} - \bar{Y}_{1.} + \bar{Y}_{..}$		48682.500	110	.52	.06
$e = (Y - \bar{Y}) - AX$	1	43150.319	200	.58	.17
"	2	35042.331	300	.65	.19
"	3	28384.646	400	.72	.19
"	4	21802.210	500	.79	.23
"	5	16230.756	600	.84	.26
"	6	11026.808	700	.89	.32
"	7	6517.490	800	.94	.41
"	8	2883.380	900	.97	.56
"	9	0	1000	1.00	1.00

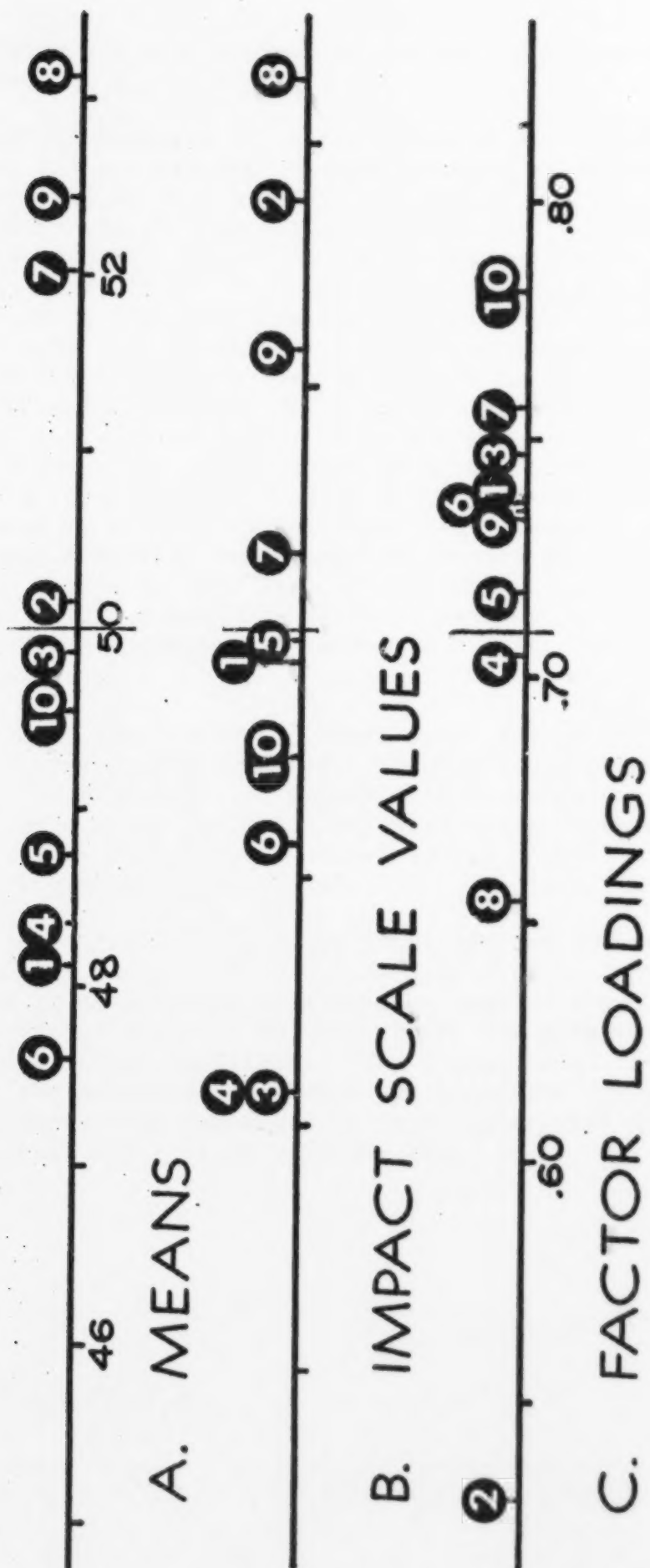


Figure 14. Three achievement scales

scale is a bipolar scale with extreme values having higher impact than middle values.

Even though the statistical significance of the impact scale was not demonstrated, the arrangement of experimental situations on the scale is of interest. We find in Figure 14 that Situations VIII and II had some opposite effect than Situations III and IV. Situation IX was something like (VIII) and (II) whereas Situation VI was something like (III) and (IV). We know that although the situation means varied somewhat, the variances were almost exactly the same (Table 22). We might have concluded that the programmed learning has a similar impact on our students as taking written notes. For certain students these conditions brought out their best performance whereas Situations III and IV did not do so. For another group of learners just the opposite was true. This dimension of impact might be interpreted as one of student organization of study. Some students do their best when obligated to plan, outline, organize. Others do their best when guided through material, making such responses as the immediate material requires. This is what Scale B in Figure 14 might have led us to conclude if more reproducible variance had been associated with it. As it stands this dimension of impact is a basis for hypotheses for our next inquiries.

To consider the similarity of impact scale values to factor loadings the data were factor analyzed. We obtained a principal axis solution for the 10x10 correlation matrix with diagonal entries of 1.00. Fifty one percent of the total variance was accounted for by the first factor, with no prominent second factor. The loadings for the first factor are shown in Figure 14.

Which of these three scales might be called our Activity-Achievement scale? Since we did not satisfy ourselves that activity was a critical variable in determining achievement, none of them. With these data the scale of means is the best basis for generalizing as to the effectiveness of study conditions. The impact model, with its scale reflecting the potential of each study situation for altering individual performance, was not found to be a substantial improvement over the conventional analysis of variance model.

BIBLIOGRAPHY

- Barclays Bank Ltd. British currency and equivalents. (free).
- Brentano, F. Von der Klassifikation der psychischen Phanomene, neue, durch Nachtrage stark vermehrte Ausgabe der betreffenden Kapitel der Psychologie vom empirischen Standpunkte, 1911.
- Briggs, L. J., Goldbeck, R. A.; Campbell, V. N.; and Nichols, D. G. Experimental results regarding form of response, size of step, and individual differences in automated programs. In Coulson (1962).
- Campbell, V. N. Adjusting self-instruction programs to individual differences: Studies of cueing, responding and bypassing. (Publ. No. AIR-C41-7/61-SR) San Mateo: American Institute for Research, 1961.
- Coulson, J. E., & Silberman, H. F. Results of an initial experiment in automated teaching. In A. A. Lumsdaine, & R. Glaser (Eds.), Teaching machines and programmed learning. Washington, D. C.: National Education Association, 1960. Pp. 452-468.
- Coulson, J. E., Ed. Programed learning and computer-based instruction. New York: Wiley, 1962.
- Cook, D. L. The Hawthorne Effect in educational research. Phi Delta Kappan, 1962, 116-122.
- Cronbach, L. J. Further evidence on response sets and test design. Educ. psychol. Measmt., 10, 1950, 3-31.
- Deese, J. The psychology of learning. Second Edition. New York: McGraw-Hill, 1958.
- Deutsche Bundesbank. Kursfahige Bundesmunzen Mai, 1959. (free).
- Dewey, J. The school and society. Chicago: University of Chicago Press, 1900.
- Dewey, J. Psychology and social practice. APA presidential address, 1899.
- Eckart, C. and Young, G. The approximation of one matrix by another of lower rank. Psychometrika, 1936, 1, 211-218.
- Evans, J. L., Glaser, R., & Homme, L. E. A preliminary investigation of variation in the properties of verbal learning sequences of the "teaching machine" type. In A. A. Lumsdaine & R. Glaser, (Eds.), Teaching machines and programed learning. Washington, D. C.: National Education Association, 1960. Pp. 446-451.

- Ferster, C. B., & Sapon, S. M. An application of recent developments in psychology to the teaching of German. Harv. educ. Rev., 1958, 28, 58-69.
- Fry, E. B. A study of teaching machine response modes. In A. A. Lumsdaine & R. Glaser, (Eds.), Teaching machines and programed learning. Washington, D. C.: National Education Association, 1960. Pp. 469-474.
- Fry, E. B. Teaching machines and programmed instruction. New York: McGraw-Hill, 1963.
- Gagne, R. M. The acquisition of knowledge. Psychol. Rev., 1962, 69, 69, 355-365.
- Gates, A. E. Recitation as a factor in memorizing. Arch. Psychol., N. Y., 1917, 6, No. 40.
- Goldbeck, R. A., & Campbell, V. N. The effects of response mode and response difficulty on programed learning. J. educ. Psychol., 1962, 53, 110-118.
- Green, B. F. and Tukey, J. W. Complex analyses of variance: general problems. Psychometrika, 25, 1960, 127-152.
- Greenhouse, S. W. and Geisser, S. On methods in the analysis of profile data. Psychometrika, 24, 1959, 95-112.
- Gulliksen, H. Linear and multidimensional scaling. Psychometrika, 1961, 26, 9-25.
- Harlow, H. F. The formation of learning sets. Psychol. Rev., 1949, 56, 51-65.
- Jain, C. and Tucker, L. R. Individual differences in the structure of color perception. Amer. J. Psychol., 1962, 75, 437-444.
- Holland, J. G. Design and use of a teaching machine program. Paper read at APA meetings, Chicago, September 1960. (mimeo)
- Horst, F. Matrix algebra for social scientists. New York: Holt, Rinehart, and Winston, 1963.
- Hottelling (1933) Analysis of a complex of statistical variables into principal components. J. educ. Psychol., 1933, 24, 417-41, 498-520.
- Hough, J. G. An analysis of the efficiency and effectiveness of selected aspects of machine instruction. J. Educ. Res., 1962, 55, 467-471.

- Jones, M. R. Current theory and research in motivation, a symposium. Lincoln: Univ. of Nebr. Press, 1953-1963.
- Kaiser, H. F. A note on Guttman's lower bounds for the number of common factors. Brit. J. stat. Psychol. XIV, 1961, p. 1.
- Keislar, E. R., & McNeil, J. D. A comparison of two response modes in an auto-instructional program with children in the primary grades. J. educ. Psychol., 1962, 53, 127-131.
- Klaus, D. English money. Pittsburgh, Pa.: American Institute for Research, 1961. (\$.50)
- Knudson, K. D. Effectiveness of overt, covert, and reading response modes in programmed instruction in English. Unpublished dissertation; George Peabody College of Education, 1962.
- Krumboltz, J. D., & Weisman, R. G. The effect of overt versus covert responding to programmed instruction on immediate and delayed retention. J. educ. Psychol., 1962, 53, 89-92.
- Lambert, P., Miller, D. M., & Wiley, D. E. Experimental folklore and experimentation: The study of programmed learning in the Wauwatosa public schools. J. educ. Res., 1962, 55, 485-494.
- Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 1953.
- Meierhenry, W. C., Ed. Learning theory and AV utilization. AV Communication Review, Supplement 4, 1961.
- Miller, G. A., Galanter, E.; and Pribram, K. H. Plans and the structure of behavior. New York: Holt, 1960.
- Myers, G. C. A study in incidental memory. Arch. Psychol., N. Y. 1913, 5, No. 26.
- Pool, R. J. Handbook of Nebraska trees. Nebraska Conservation Bulletin No. 32. Lincoln, Nebraska: Univ. of Nebraska Conservation and Survey Division, 1951. (\$.50)
- Porter, D. Some effects of year long teaching machine instruction. In E. H. Galanter (Ed.), Automatic teaching: The state of the art. New York: Wiley, 1959. Pp. 85-90.
- Postman, L. and Senders, V. L. Incidental learning and generality of set. J. exp. Psychol., 1946, 36, 153-165.
- Postman, L. The history and present status of the law of effect. Psychol. Bull. 1947, 44, 489-563.

- Pressey, S. L. Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. J. Psychol., 1950, 29, 417-447.
- Resnick, L. B. Programed instruction and the teaching of complex intellectual skills: problems and prospects. Harvard educ. Rev., 1963, 33, 439-471.
- Rigney, J. W. Potential uses of computers as teaching machines. In Coulson (1962).
- Robinson, F. P. Effective study. New York: Harper, 1946.
- Roe, Arnold, Research in programed learning. From J. E. Coulson, op. cit., (1962).
- Silberman, H. F. Characteristics of some recent studies of instruction methods. In Coulson (1962).
- Skinner, B. F. The science of learning and the art of teaching. Harvard educ. Rev., 24, 1954, 86-97.
- Skinner, B. F. Teaching machines. Science, 1958, 128, 969-977.
- Stake, R. E. The teaching machine: tool of the future of passing fancy. Phi Delta Kappan. XLIV, 1963, 247-249.
- Stolurow, L. M. Teaching by machine. OE-34010. Washington, D. C.: U. S. Dept. of Health, Education, and Welfare. 1961.
- Stone, G. R. A note on Postman's review of the literature on the law of effect. Psychol. Bull., 1948, 45, 151-160.
- Thorndike, E. L. The law of effect. Amer. J. Psychol., 1927, 39, 212-222.
- Torgerson, W. S. Theory and methods of scaling. New York: Wiley, 1958.
- Tucker, L. R. and Messick, S. An individual differences model for multidimensional scaling. Psychometrika, 1963, 28, 333-367.
- Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.
- Zeigarnik, B. Veber das Behalten von erledigten and unerledigten Handlungen. Psychol. Forsch., 1927, 9, 1-85.